REMARKS

Favorable reconsideration, reexamination, and allowance of the present patent application are respectfully requested in view of the following remarks.

Rejection under 35 U.S.C. § 112, first paragraph

In the Office Action, beginning at page 2, Claims 1, 6, 7, and 10-12 were rejected under 35 U.S.C. § 112, first paragraph, as failing to comply with the written description requirement. Applicant respectfully requests reconsideration of this rejection.

First, to address the comments in the Office Action (page 3) concerning the alleged lack of description of the genus of *rmf* genes, Applicants submit an alignment of *E.coli rmf* genes from varous strains of *E.coli*. The Office action states that the claims are directed to a genus of *rmf* gene; however, the attached alignment shows that the genus of known *E.coli rmf* genes contains species which are identical (see attached Exhibit A), except for one strain, which differs in one amino acid. Therefore, the claimed expression "endogenous *Escherichia coli* gene encoding the RMF protein" clearly represents a genus which is known in the art, and includes little variation. Therefore, the genus is fully described.

Secondly, Applicants respectfully assert that the citation in the Office to the Federal Circuit decision in *Eli Lilly* is inapplicable. The passage from the Federal Register cited by the Examiner states that "a chemical compound's name does not necessarily convey a written description of the named chemical compound" (emphasis added). But clearly, the court contemplated situations when a name IS sufficient to satisfy written description. Consistent with their stance in *Eli Lilly* is the Court's decision in *Capon v. Eshhar* (03-1480, -1481) (Exhibit B), where the Court noted strongly that recitation of structure, and in this case, a sequence, in the specification is NOT necessary when the sequence and/or structure is known in the art. Clearly, the Court does not want to require applicants to recite prior art sequences in the specification in order to

adequately describe an invention using a known sequence, which is the main tenant of *Capon*.

The Examiner has stated that "the specification does not describe and define any structural features and nucleotide sequences commonly possessed by the genus." However, as the attached alignment shows, the genus of *E.coli rmf* genes were known in the art. The Court has consistently held that recitation of that which is known in the prior art is not required in the specification, and is preferably not recited.

These principles clearly apply to the facts of the instant application. As stated above, the *rmf* gene from *E. coli* has been described in the prior art, including its structure and function. As the attached alignment in Exhibit A shows, the genus of *E.coli rmf* genes actually includes little to no variation. Furthermore, the techniques used for disruption of the gene are well-known. Specifically, the Example 2 in the specification cites to the article by Link et al. (see Exhibit C), and which is attached hereto for the Examiner's convenience. Link et al. describes the disruption of a gene gene from *E. coli* by crossover PCR, showing that such procedures are well-known in the art.

In regard to the assertion in the Office Action that other types of amino acids, such as aliphatic, aromatic, hydroxylic, etc., are not shown as being increased by the methods of the invention, Applicants respectfully submit a declaration under 37 C.F.R. §1.132 by Dr. Imaizumi, on the the inventors, which describes experimental results showing that inactivation of the RMF protein is also effective to increase production of L-tryptophan. Therefore, it has been demonstrated that production of L-lysine (example 2 in the specification), L-glutamic acid (Rule 132 declaration submitted on March 28, 2005), and L-tryptophan (declaration submitted herewith) can be increased by inactivation of the RMF protein. These exemplified amino acids are very diverse, and represent very different structures, demonstrating a wide range of the claimed genus production method. Specifically, such data exemplifies production of basic, acidic, and aromatic amino acids, demonstrating the application of the method to different kinds of

amino acids. Therefore, production to any amino acid is sufficiently supported and adequately described.

The claims have been additionally rejected as gene elements, including regulatory elements, expression control sequences, and untranslated regions are allegedly not described. The Office Action points to differences in structural elements of genes which mediate expression of a liver protein versus the same protein in brain. Such a comparison is entirely unfounded, since the method of the present invention occurs in bacteria, a single cell, simple, well-understood organism. Bacterial gene elements and manipulation of such were well-known before the priority date of the present application. Again, description of that which is well-known in the art is not required, and is preferably omitted, from a specification. Therefore, the claims, including the methods for manipulating expression by altering endogenous *E. coli* expression elements of a known gene, are sufficiently described.

For at least the foregoing reasons, Applicant respectfully submits that Claims 1, 6, 7, and 10-12 fully comply with 35 U.S.C. § 112, first paragraph, regarding written desription and therefore respectfully requests withdrawal of the rejection thereof under 35 U.S.C. § 112.

In the Office Action, beginning at page 4, Claims 7 and 12 were rejected under 35 U.S.C. § 112, first paragraph, allegedly failed to comply with the enablement requirement. Applicant respectfully requests reconsideration of this rejection.

Similar to the above rejection, the basis for this rejection appears to be the lack of a disclosure of the specific nucleotide sequence on the rmf gene which is obtained by crossover PCR. However, as stated above, and shown in the attached Exhibits, the nucleotide sequence of the *rmf* gene and the crossover PCR method used to disrupt this gene as shown in example 2 in the specification have been known since well before the priority date of the application. The Link et al. article (Exhibit C) clearly shows the use

of cross-over PCR for disruption of the *E.coli* genome by crossover PCR, and that such techniques were well-known in the art. Therefore, the WC196Δrmf strain can be prepared from the deposited strain WC196 (AJ13069) according to the description of the specification and in light of that which was known in the art. Therefore, since the WC196Δrmf strain was readily obtainable by the methods set forth in the specification combined with the knowledge in the art, a deposit of the strains is unnecessary.

As for the assertion in the Office Action that the specification does not disclose the specific nucleotide sequence of the inactivated gene, applicants assert that the specification does disclose the sequences of the four primers. As is shown in the attached alignment, and in the prior art (see Yamagishi et al., *EMBO J.* (1993) 12:625-630, cited in the specification by applicants and throughout prosecution by the Examiner), the entire sequence of the *rmf* gene was known. Clearly such knowledge in the prior art combined with the primer sequences in the specification, provides a clear description of the inactivated gene since one of skill in the art could easily determine the complementary regions based up on the primers, and determine the structure of the inactivated gene following the well-known prior technique of cross-over PCR.

For at least the foregoing reasons, Applicant respectfully submits that Claims 7 and 12 comply with 35 U.S.C. § 112, first paragraph, enablement, and therefore respectfully requests withdrawal of the rejection thereof under 35 U.S.C. § 112.

Conclusion

For at least the foregoing reasons, Applicant respectfully submits that the present patent application is in condition for allowance. An early indication of the allowability of the present patent application is therefore respectfully solicited.

If Examiner Fronda believes that a telephone conference with the undersigned would expedite passage of the present patent application to issue, he is invited to call on the number below.

It is not believed that extensions of time are required, beyond those that may otherwise be provided for in accompanying documents. However, if additional extensions of time are necessary to prevent abandonment of this application, then such extensions of time are hereby petitioned under 37 C.F.R. § 1.136(a), and the undersigned respectfully requests that any such necessary fees be charged to our deposit account 50-2821.

Respectfully submitted,

By: _____

Shelly Guest Cermak
Registration No. 39,571

U.S. P.T.O. Customer No. 38108
Cermak & Kenealy, LLP
515 E. Braddock Road, Suite B
Alexandria, VA 22314
703.778.6608

Date: May 5, 2006

EXHIBIT
A

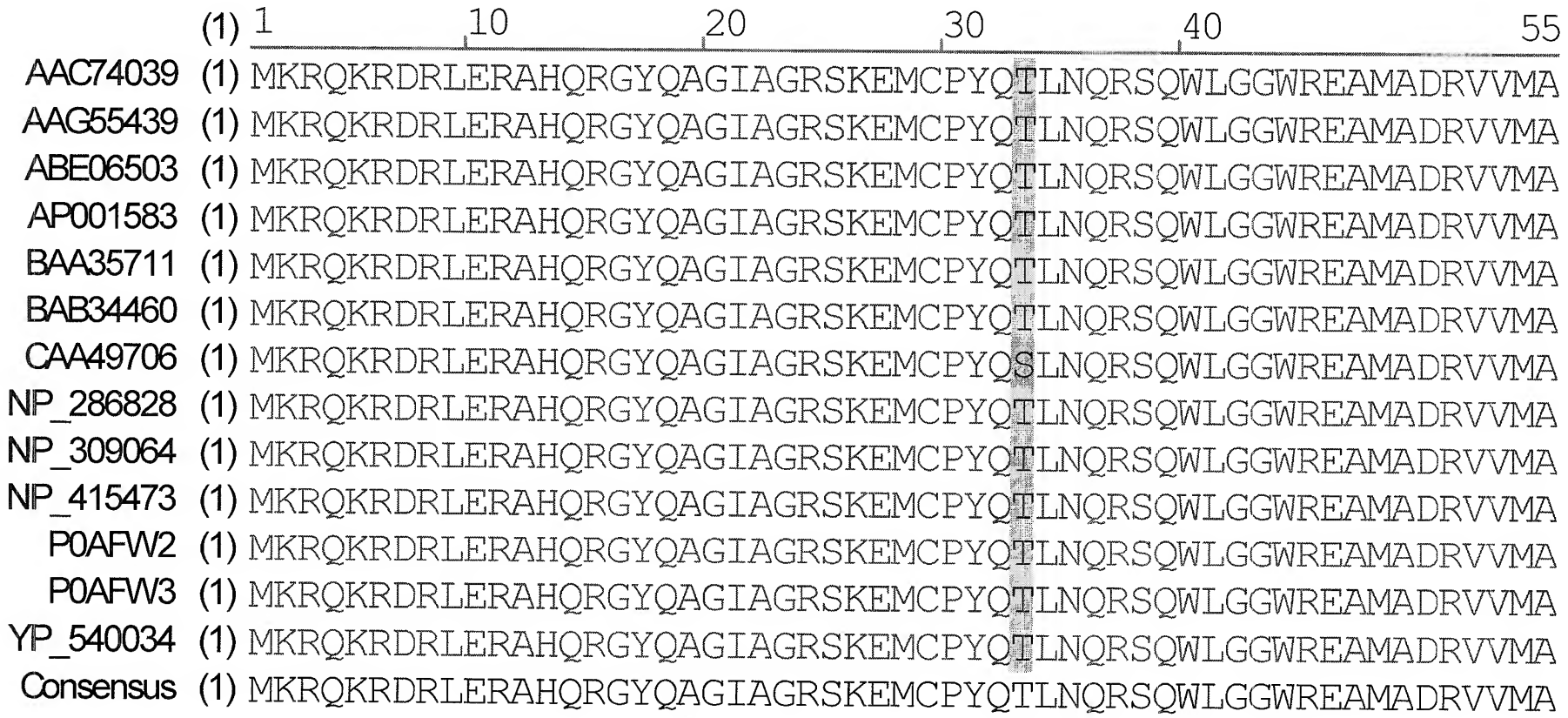
(1) 1 10 20 30 40 55
AAC74039 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
AAG55439 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
ABE06503 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
AP001583 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
BAA35711 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
BAB34460 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
CAA49706 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQSLNQRSQWLGGWREAMADRVVMA
NP_286828 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
NP_309064 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
NP_415473 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
P0AFW2 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
P0AFW3 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
YP_540034 (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA
Consensus (1) MKRQKRDRLERAHQRGYQAGIAGRSKEMCPYQTLNQRSQWLGGWREAMADRVVMA

EXHIBIT
B

United States Court of Appeals for the Federal Circuit

03-1480, -1481
(Interference No. 103,887)

DANIEL J. CAPON, ARTHUR WEISS, BRIAN A. IRVING, MARGO R. ROBERTS, and KRISZTINA ZSEBO,

Appellants,

v.

ZELIG ESHHAR, DANIEL SCHINDLER, TOVA WAKS, and GIDEON GROSS,

Cross-Appellants,

v.

JON DUDAS, Director of the Patent and Trademark Office,

Intervenor.

Steven B. Kelber, Piper Rudnick, LLP, of Washington, DC, argued for appellants.

Roger L. Browdy, Browdy and Neimark, P.L.L.C., of Washington, DC, argued for cross-appellants.

Mary L. Kelly, Associate Solicitor, Office of the Solicitor, United States Patent and Trademark Office, of Arlington, Virginia, argued for intervenor. With her on the brief were John M. Whealan, Solicitor and Stephen Walsh, Associate Solicitor.

Appealed from: United States Patent and Trademark Office Board of Patent Appeals and Interferences

United States Court of Appeals for the Federal Circuit

03-1480, -1481
(Interference No. 103,887)

DANIEL J. CAPON, ARTHUR WEISS, BRIAN A. IRVING,
MARGO R. ROBERTS, and KRISZTINA ZSEBO,
Appellants,

v.

ZELIG ESHHAR, DANIEL SCHINDLER, TOVA WAKS,
and GIDEON GROSS,
Cross-Appellants,

v.

JON DUDAS,
Director of the Patent and Trademark Office,
Intervenor.

DECIDED: August 12, 2005

Before NEWMAN, MAYER,* and GAJARSA, Circuit Judges.

NEWMAN, Circuit Judge.

Both of the parties to a patent interference proceeding have appealed the decision of the Board of Patent Appeals and Interferences of the United States Patent and Trademark Office, wherein the Board held that the specification of neither party met the written description requirement of the patent statute. Capon v. Eshhar, Interf. No. 103,887

---

* Haldane Robert Mayer vacated the position of Chief Judge on December 24, 2004.

(Bd. Pat. App. & Interf. Mar. 26, 2003). The Board dissolved the interference and cancelled all of the claims of both parties corresponding to the interference count. With this ruling, the Board terminated the proceeding and did not reach the question of priority of invention. We conclude that the Board erred in its application of the law of written description. The decision is vacated and the case is remanded to the Board for further proceedings.

BACKGROUND

Daniel J. Capon, Arthur Weiss, Brian A. Irving, Margo R. Roberts, and Krisztina Zsebo (collectively "Capon") and Zelig Eshhar, Daniel Schindler, Tova Waks, and Gideon Gross (collectively "Eshhar") were the parties to an interference proceeding between Capon's United States Patent No. 6,407,221 ("the '221 patent") entitled "Chimeric Chains for Receptor-Associated Signal Transduction Pathways" and Eshhar's patent application Serial No. 08/084,994 ("the '994 application") entitled "Chimeric Receptor Genes and Cells Transformed Therewith." Capon's Patent No. 5,359,046 ("the '046 patent"), parent of the '221 patent, was also included in the interference but was held expired for non-payment of a maintenance fee. The PTO included the '046 patent in its decision and in its argument of this appeal.[1]

A patent interference is an administrative proceeding pursuant to 35 U.S.C. §§102(g) and 135(a), conducted for the purpose of determining which of competing applicants is the first inventor of common subject matter. An interference is instituted after the separate

---

1 Although Capon is designated as appellant and Eshhar as cross-appellant, both appealed the Board's decision. See Fed. R. App. P. 28(h). The Director of the PTO intervened to support the Board, and has fully participated in this appeal.

patent applications have been examined and found to contain patentable subject matter. Capon's patents had been examined and had issued before this interference was instituted, and Eshhar's application had been examined and allowed but a patent had not yet issued.

During an interference proceeding the Board is authorized to determine not only priority of invention but also to redetermine patentability. 35 U.S.C. §6(b). The question of patentability of the claims of both parties was raised *sua sponte* by an administrative patent judge during the preliminary proceedings. Thereafter the Board conducted an *inter partes* proceeding limited to this question, receiving evidence and argument. The Board then invalidated all of the claims that had been designated as corresponding to the count of the interference, viz., all of the claims of the Capon '221 patent, claims 5-8 of the Capon '046 patent, and claims 1-7, 9-20, and 23 of the Eshhar '994 application.

In accordance with the Administrative Procedure Act, the law as interpreted and applied by the agency receives plenary review on appeal, and the agency's factual findings are reviewed to determine whether they were arbitrary, capricious, or unsupported by substantial evidence in the administrative record. See 5 U.S.C. §706(2); Dickinson v. Zurko, 527 U.S. 150, 164-65 (1999); In re Gartside, 203 F.3d 1305, 1315 (Fed. Cir. 2000).

The Invention

A chimeric gene is an artificial gene that combines segments of DNA in a way that does not occur in nature. The '221 patent and '994 application are directed to the production of chimeric genes designed to enhance the immune response by providing cells with specific cell-surface antibodies in a form that can penetrate diseased sites, such as solid tumors, that were not previously reachable. The parties explain that their invention is

a way of endowing immune cells with antibody-type specificity, by combining known antigen-binding-domain producing DNA and known lymphocyte-receptor-protein producing DNA into a unitary gene that can express a unitary polypeptide chain. Eshhar summarized the problem to which the invention is directed:

> Antigen-specific effector lymphocytes, such as tumor-specific T cells, are very rare, individual-specific, limited in their recognition spectrum and difficult to obtain against most malignancies. Antibodies, on the other hand, are readily obtainable, more easily derived, have wider spectrum and are not individual-specific. The major problem of applying specific antibodies for cancer immunotherapy lies in the inability of sufficient amounts of monoclonal antibodies (mAb) to reach large areas within solid tumors.

Technical Paper Explaining Eshhar's Invention, at 6.

The inventions of Capon and Eshhar are the chimeric DNA that encodes single-chain chimeric proteins for expression on the surface of cells of the immune system, plus expression vectors and cells transformed by the chimeric DNA. The experts for both parties explain that the invention combines selected DNA segments that are both endogenous and nonendogenous to a cell of the immune system, whereby the nonendogenous segment encodes the single-chain variable ("scFv") domain of an antibody, and the endogenous segment encodes cytoplasmic, transmembrane, and extracellular domains of a lymphocyte signaling protein. They explain that the scFv domain combines the heavy and light variable ("Fv") domains of a natural antibody, and thus has the same specificity as a natural antibody. Linking this single chain domain to a lymphocyte signaling protein creates a chimeric scFv-receptor ("scFvR") gene which, upon transfection into a cell of the immune system, combines the specificity of an antibody with the tissue penetration, cytokine production, and target-cell destruction capability of a lymphocyte.

The parties point to the therapeutic potential if tumors can be infiltrated with specifically designed immune cells of appropriate anti-tumor specificity.

The Eshhar Claims

The Board held unpatentable the following claims of Eshhar's '994 application; these were all of the '994 claims that had been designated as corresponding to the count of the interference. Eshhar's claim 1 was the designated count.

> 1. A chimeric gene comprising
>
> a first gene segment encoding a single-chain Fv domain (scFv) of a specific antibody and
>
> a second gene segment encoding partially or entirely the transmembrane and cytoplasmic, and optionally the extracellular, domains of an endogenous protein
>
> wherein said endogenous protein is expressed on the surface of cells of the immune system and triggers activation and/or proliferation of said cells,
>
> which chimeric gene, upon transfection to said cells of the immune system, expresses said scFv domain and said domains of said endogenous protein in one single chain on the surface of the transfected cells such that the transfected cells are triggered to activate and/or proliferate and have MHC nonrestricted antibody-type specificity when said expressed scFV domain binds to its antigen.
>
> 2. A chimeric gene according to claim 1 wherein the second gene segment further comprises partially or entirely the extracellular domain of said endogenous protein.
>
> 3. A chimeric gene according to claim 1 wherein the first gene segment encodes the scFv domain of an antibody against tumor cells.
>
> 4. A chimeric gene according to claim 1 wherein the first gene segment encodes the scFv domain of an antibody against virus infected cells.
>
> 5. A chimeric gene according to claim 4 wherein the virus is HIV.
>
> 6. A chimeric gene according to claim 1 wherein the second gene segment encodes a lymphocyte receptor chain.
>
> 7. A chimeric gene according to claim 6 wherein the second gene segment encodes a chain of the T cell receptor.

9. A chimeric gene according to claim 7 wherein the second gene segment encodes the α, β, γ, or δ chain of the antigen-specific T cell receptor.

10. A chimeric gene according to claim 1 wherein the second gene segment encodes a polypeptide of the TCR/CD3 complex.

11. A chimeric gene according to claim 10 wherein the second gene segment encodes the zeta or eta isoform chain.

12. A chimeric gene according to claim 1 wherein the second gene segment encodes a subunit of the Fc receptor or IL-2 receptor.

13. A chimeric gene according to claim 12 wherein the second gene segment encodes a common subunit of IgE and IgG binding Fc receptors.

14. A chimeric gene according to claim 13 wherein said subunit is the gamma subunit.

15. A chimeric gene according to claim 13 wherein the second gene segment encodes the CD16α chain of the FcγRIII or FcγRII.

16. A chimeric gene according to claim 12 wherein the second gene segment encodes the α or β subunit of the IL-2 receptor.

17. An expression vector comprising a chimeric gene according to claim 1.

18. A cell of the immune system endowed with antibody specificity transformed with an expression vector according to claim 17.

19. A cell of the immune system endowed with antibody specificity comprising a chimeric gene according to claim 1.

20. A cell if the immune system according to claim 19 selected from the group consisting of a natural killer cell, a lymphokine activated killer cell, a cytotoxic T cell, a helper T cell and a subtype thereof.

23. A chimeric gene according to claim 1 wherein said endogenous protein is a lymphocyte receptor chain, a polypeptide of the TCR/CD3 complex, or a subunit of the Fc or IL-2 receptor.

The Board did not discuss the claims separately, and held that the specification failed to satisfy the written description requirement as to all of these claims.

The Capon Claims

Claims 1-10, all of the claims of the '221 patent, were held unpatentable on written description grounds. Claims 1-6 are directed to the chimeric DNA, claims 7, 8, and 10 to the corresponding cell comprising the DNA, and claim 9 to the chimeric protein:

> 1. A chimeric DNA encoding a membrane bound protein, said chimeric DNA comprising in reading frame:
>
> DNA encoding a signal sequence which directs said membrane bound protein to the surface membrane;
>
> DNA encoding a non-MHC restricted extracellular binding domain which is obtained from a single chain antibody that binds specifically to at least one ligand, wherein said at least one ligand is a protein on the surface of a cell or a viral protein;
>
> DNA encoding a transmembrane domain which is obtained from a protein selected from the group consisting of CD4, CD8, immunoglobulin, the CD3 zeta chain, the CD3 gamma chain, the CD3 delta chain and the CD3 epsilon chain; and
>
> DNA encoding a cytoplasmic signal-transducing domain of a protein that activates an intracellular messenger system which is obtained from CD3 zeta,
>
> wherein said extracellular domain and said cytoplasmic domain are not naturally joined together, and said cytoplasmic domain is not naturally joined to an extracellular ligand-binding domain, and when said chimeric DNA is expressed as a membrane bound protein in a host cell under conditions suitable for expression, said membrane bound protein initiates signaling in said host cell when said extracellular domain binds said at least one ligand.
>
> 2. The DNA of claim 1, wherein said single-chain antibody recognizes an antigen selected from the group consisting of viral antigens and tumor cell associated antigens.
>
> 3. The DNA of claim 2 wherein said single-chain antibody is specific for the HIV env glycoprotein.
>
> 4. The DNA of claim 1, wherein said transmembrane domain is naturally joined to said cytoplasmic domain.
>
> 5. An expression cassette comprising a transcriptional initiation region, the DNA of claim 1 under the transcriptional control of said transcriptional initiation region, and a transcriptional termination region.

6. A retroviral RNA or DNA construct comprising the expression cassette of claim 5.

7. A cell comprising the DNA of claim 1.

8. The cell of claim 7, wherein said cell is a human cell.

9. A chimeric protein comprising in the N-terminal to C-terminal direction:

a non-MHC restricted extracellular binding domain which is obtained from a single chain antibody that binds specifically to at least one ligand, wherein said at least one ligand is a protein on the surface of a cell or a viral protein;

a transmembrane domain which is obtained from a protein selected from the group consisting CD4, CD8, immunoglobulin, the CD3 zeta chain, the CD3 gamma chain, the CD3 delta chain and the CD3 epsilon chain; and

a cytoplasmic signal-transducing domain of a protein that activates an intracellular messenger system which is obtained from CD3 zeta,

wherein said extracellular domain and said cytoplasmic domain are not naturally joined together, and said cytoplasmic domain is not naturally joined to an extracellular ligand-binding domain, and when said chimeric protein is expressed as a membrane bound protein in a host cell under conditions suitable for expression, said membrane bound protein initiates signaling in said host cell when said extracellular domain binds said at least one ligand.

10. A mammalian cell comprising as a surface membrane protein, the protein of claim 9.

In addition, claims 5, 6, 7, and 8 of Capon's '046 patent were held unpatentable. These claims are directed to chimeric DNA sequences where the encoded extracellular domain is a single-chain antibody containing ligand binding activity.

The Board Decision

The Board presumed enablement by the specifications of the '221 patent and '994 application of the full scope of their claims, and based its decision solely on the ground of

failure of written description. The Board held that neither party's specification provides the requisite description of the full scope of the chimeric DNA or encoded proteins, by reference to knowledge in the art of the "structure, formula, chemical name, or physical properties" of the DNA or the proteins. In the Board's words:

> We are led by controlling precedent to understand that the full scope of novel chimeric DNA the parties claim is not described in their specifications under 35 U.S.C. §112, first paragraph, by reference to contemporary and/or prior knowledge in the art of the structure, formula, chemical name, or physical properties of many protein domains, and/or DNA sequences which encode many protein domains, which comprise single-chain proteins and/or DNA constructs made in accordance with the plans, schemes, and examples thereof the parties disclose.

Bd. op. at 4. As controlling precedent the Board cited Regents of the University of California v. Eli Lilly & Co., 119 F.3d 1559 (Fed. Cir. 1997); Fiers v. Revel Co., 984 F.2d 1164 (Fed. Cir. 1993); Amgen, Inc. v. Chugai Pharmaceutical Co., 927 F.2d 1200 (Fed. Cir. 1991); and Enzo Biochem, Inc. v. Gen-Probe, Inc., 296 F.3d 1316 (Fed. Cir. 2002). The Board summarized its holding as follows:

> Here, both Eshhar and Capon claim novel genetic material described in terms of the functional characteristics of the protein it encodes. Their specifications do not satisfy the written description requirement because persons having ordinary skill in the art would not have been able to visualize and recognize the identity of the claimed genetic material without considering additional knowledge in the art, performing additional experimentation, and testing to confirm results.

Bd. op. at 89.

DISCUSSION

Eshhar and Capon challenge both the Board's interpretation of precedent and the Board's ruling that their descriptions are inadequate. Both parties explain that their

chimeric genes are produced by selecting and combining known heavy- and light-chain immune-related DNA segments, using known DNA-linking procedures. The specifications of both parties describe procedures for identifying and obtaining the desired immune-related DNA segments and linking them into the desired chimeric genes. Both parties point to their specific examples of chimeric DNA prepared using identified known procedures, along with citation to the scientific literature as to every step of the preparative method.

The parties presented expert witnesses who placed the invention in the context of prior knowledge and explained how the descriptive text would be understood by persons of skill in the field of the invention. The witnesses explained that the principle of forming chimeric genes from selected segments of DNA was known, as well as their methods of identifying, selecting, and combining the desired segments of DNA. Dr. Eshhar presented an expert statement wherein he explained that the prior art contains extensive knowledge of the nucleotide structure of the various immune-related segments of DNA; he stated that over 785 mouse antibody DNA light chains and 1,327 mouse antibody DNA heavy chains were known and published as early as 1991. Similarly Capon's expert Dr. Desiderio discussed the prior art, also citing scientific literature:

> The linker sequences disclosed in the '221 patent (col. 24, lines 4 and 43) used to artificially join a heavy and light chain nucleic acid sequence and permit functional association of the two ligand binding regions were published by 1990, as were the methods for obtaining the mature sequences of the desired heavy and light chains for constructing a SAb (Exhibit 47, Batra et al., J., Biol. Chem., 1990; Exhibit 48, Bird et al., Science, 1988; Exhibit 50, Huston et al., PNAS, 1988; Exhibit 51, Chaudhary, PNAS, 1990, Exhibit 56, Morrison et al., Science, 1985; Exhibit 53, Sharon et al., Nature 1984).

Desiderio declaration at 4 ¶11.

Both parties stated that persons experienced in this field would readily know the structure of a chimeric gene made of a first segment of DNA encoding the single-chain variable region of an antibody, and a second segment of DNA encoding an endogenous protein. They testified that re-analysis to confirm these structures would not be needed in order to know the DNA structure of the chimeric gene, and that the Board's requirement that the specification must reproduce the "structure, formula, chemical name, or physical properties" of these DNA combinations had been overtaken by the state of the science. They stated that where the structure and properties of the DNA components were known, reanalysis was not required.

Eshhar's specification contains the nucleotide sequences of sixteen different receptor primers and four different scFv primers from which chimeric genes encoding scFvR may be obtained, while Capon's specification cites literature sources of such information. Eshhar's specification shows the production of chimeric genes encoding scFvR using primers, as listed in Eshhar's Table I. Capon stated that natural genes are isolated and joined using conventional methods, such as the polymerase chain reaction or cloning by primer repair. Capon, like Eshhar, discussed various known procedures for identifying, obtaining, and linking DNA segments, accompanied by experimental examples. The Board did not dispute that persons in this field of science could determine the structure or formula of the linked DNA from the known structure or formula of the components.

The Board stated that "controlling precedent" required inclusion in the specification of the complete nucleotide sequence of "at least one" chimeric gene. Bd. op. at 4. The Board also objected that the claims were broader than the specific examples. Eshhar and

Capon each responds by pointing to the scientific completeness and depth of their descriptive texts, as well as to their illustrative examples. The Board did not relate any of the claims, broad or narrow, to the examples, but invalidated all of the claims without analysis of their scope and the relation of claim scope to the details of the specifications.

Eshhar and Capon both argue that they have set forth an invention whose scope is fully and fairly described, for the nucleotide sequences of the DNA in chimeric combination is readily understood to contain the nucleotide sequences of the DNA components. Eshhar points to the general and specific description in his specification of known immune-related DNA segments, including the examples of their linking. Capon points similarly to his description of selecting DNA segments that are known to express immune-related proteins, and stresses the existing knowledge of these segments and their nucleotide sequences, as well as the known procedures for selecting and combining DNA segments, as cited in the specification.

Both parties argue that the Board misconstrued precedent, and that precedent does not establish a *per se* rule requiring nucleotide-by-nucleotide re-analysis when the structure of the component DNA segments is already known, or readily determined by known procedures.

The Statutory Requirement

The required content of the patent specification is set forth in Section 112 of Title 35:

> §112 ¶1. The specification shall contain a written description of the invention, and of the manner and process of making and using it, in such full,

> clear, concise, and exact terms as to enable any person skilled in the art to which it pertains, or with which it is most nearly connected, to make and use the same, and shall set forth the best mode contemplated by the inventor of carrying out his invention.

The "written description" requirement implements the principle that a patent must describe the technology that is sought to be patented; the requirement serves both to satisfy the inventor's obligation to disclose the technologic knowledge upon which the patent is based, and to demonstrate that the patentee was in possession of the invention that is claimed. See Enzo Biochem, 296 F.3d at 1330 (the written description requirement "is the quid pro quo of the patent system; the public must receive meaningful disclosure in exchange for being excluded from practicing the invention for a limited period of time"); Reiffin v. Microsoft Corp., 214 F.3d 1342, 1345-46 (Fed. Cir. 2000) (the purpose of the written description requirement "is to ensure that the scope of the right to exclude . . . does not overreach the scope of the inventor's contribution to the field of art as described in the patent specification"); In re Barker, 559 F.2d 588, 592 n.4 (CCPA 1977) (the goal of the written description requirement is "to clearly convey the information that an applicant has invented the subject matter which is claimed"). The written description requirement thus satisfies the policy premises of the law, whereby the inventor's technical/scientific advance is added to the body of knowledge, as consideration for the grant of patent exclusivity.

The descriptive text needed to meet these requirements varies with the nature and scope of the invention at issue, and with the scientific and technologic knowledge already in existence. The law must be applied to each invention that enters the patent process, for each patented advance is novel in relation to the state of the science. Since the law is applied to each invention in view of the state of relevant knowledge, its application will vary

with differences in the state of knowledge in the field and differences in the predictability of the science.

For the chimeric genes of the Capon and Eshhar inventions, the law must take cognizance of the scientific facts. The Board erred in refusing to consider the state of the scientific knowledge, as explained by both parties, and in declining to consider the separate scope of each of the claims. None of the cases to which the Board attributes the requirement of total DNA re-analysis, i.e., Regents v. Lilly, Fiers v. Revel, Amgen, or Enzo Biochem, require a re-description of what was already known. In Lilly, 119 F.3d at 1567, the cDNA for human insulin had never been characterized. Similarly in Fiers, 984 F.2d at 1171, much of the DNA sought to be claimed was of unknown structure, whereby this court viewed the breadth of the claims as embracing a "wish" or research "plan." In Amgen, 927 F.2d at 1206, the court explained that a novel gene was not adequately characterized by its biological function alone because such a description would represent a mere "wish to know the identity" of the novel material. In Enzo Biochem, 296 F.3d at 1326, this court reaffirmed that deposit of a physical sample may replace words when description is beyond present scientific capability. In Amgen Inc. v. Hoechst Marion Roussel, Inc., 314 F.3d 1313, 1332 (Fed. Cir. 2003) the court explained further that the written description requirement may be satisfied "if in the knowledge of the art the disclosed function is sufficiently correlated to a particular, known structure." These evolving principles were applied in Noelle v. Lederman, 355 F.3d 1343, 1349 (Fed. Cir. 2004), where the court affirmed that the human antibody there at issue was not adequately described by the structure and function of the mouse antigen; and in University of Rochester v. G.D. Searle & Co., 358 F.3d 916, 925-26 (Fed.

Cir. 2004), where the court affirmed that the description of the COX-2 enzyme did not serve to describe unknown compounds capable of selectively inhibiting the enzyme.

The "written description" requirement must be applied in the context of the particular invention and the state of the knowledge. The Board's rule that the nucleotide sequences of the chimeric genes must be fully presented, although the nucleotide sequences of the component DNA are known, is an inappropriate generalization. When the prior art includes the nucleotide information, precedent does not set a *per se* rule that the information must be determined afresh. Both parties state that a person experienced in the field of this invention would know that these known DNA segments would retain their DNA sequences when linked by known methods. Both parties explain that their invention is not in discovering which DNA segments are related to the immune response, for that is in the prior art, but in the novel combination of the DNA segments to achieve a novel result.

The "written description" requirement states that the patentee must describe the invention; it does not state that every invention must be described in the same way. As each field evolves, the balance also evolves between what is known and what is added by each inventive contribution. Both Eshhar and Capon explain that this invention does not concern the discovery of gene function or structure, as in Lilly. The chimeric genes here at issue are prepared from known DNA sequences of known function. The Board's requirement that these sequences must be analyzed and reported in the specification does not add descriptive substance. The Board erred in holding that the specifications do not meet the written description requirement because they do not reiterate the structure or formula or chemical name for the nucleotide sequences of the claimed chimeric genes.

Claim Scope

There remains the question of whether the specifications adequately support the breadth of all of the claims that are presented. The Director argues that it cannot be known whether all of the permutations and combinations covered by the claims will be effective for the intended purpose, and that the claims are too broad because they may include inoperative species. The inventors say that they have provided an adequate description and exemplification of their invention as would be understood by persons in the field of the invention. They state that biological properties typically vary, and that their specifications provide for evaluation of the effectiveness of their chimeric combinations.

It is well recognized that in the "unpredictable" fields of science, it is appropriate to recognize the variability in the science in determining the scope of the coverage to which the inventor is entitled. Such a decision usually focuses on the exemplification in the specification. See, e.g., Enzo Biochem, 296 F.3d at 1327-28 (remanding for district court to determine "[w]hether the disclosure provided by the three deposits in this case, coupled with the skill of the art, describes the genera of claims 1-3 and 5"); Lilly, 119 F.3d at 1569 (genus not described where "a representative number of cDNAs, defined by nucleotide sequence, falling within the scope of the genus" had not been provided); In re Gostelli, 872 F.2d 1008, 1012 (Fed. Cir. 1989) (two chemical compounds were insufficient description of subgenus); In re Smith, 458 F.2d 1389, 1394-95 (CCPA 1972) (disclosure of genus and one species was not sufficient description of intermediate subgenus); In re Grimme, 274 F.2d 949, 952 (CCPA 1960) (disclosure of single example and statement of scope sufficient disclosure of subgenus).

Precedent illustrates that the determination of what is needed to support generic claims to biological subject matter depends on a variety of factors, such as the existing

knowledge in the particular field, the extent and content of the prior art, the maturity of the science or technology, the predictability of the aspect at issue, and other considerations appropriate to the subject matter. See, e.g., In re Wallach, 378 F.3d 1330, 1333-34 (Fed. Cir. 2004) (an amino acid sequence supports "the entire genus of DNA sequences" that can encode the amino acid sequence because "the state of the art has developed" such that it is a routine matter to convert one to the other); University of Rochester, 358 F.3d at 925 (considering whether the patent disclosed the compounds necessary to practice the claimed method, given the state of technology); Singh v. Brake, 317 F.3d 1334, 1343 (Fed. Cir. 2002) (affirming adequacy of disclosure by distinguishing precedent in which the selection of a particular species within the claimed genus had involved "highly unpredictable results").

It is not necessary that every permutation within a generally operable invention be effective in order for an inventor to obtain a generic claim, provided that the effect is sufficiently demonstrated to characterize a generic invention. See In re Angstadt, 537 F.2d 498, 504 (CCPA 1976) ("The examples, both operative and inoperative, are the best guidance this art permits, as far as we can conclude from the record"). While the Board is correct that a generic invention requires adequate support, the sufficiency of the support must be determined in the particular case. Both Eshhar and Capon present not only general teachings of how to select and recombine the DNA, but also specific examples of the production of specified chimeric genes. For example, Eshhar points out that in Example 1 of his specification the FcRγ chain was used, which chain was amplified from a human cDNA clone, using the procedure of Kuster, H. et al., J. Biol. Chem., 265:6448-6451 (1990), which is cited in the specification and reports the complete sequence of the FcRγ

chain. Eshhar's Example 1 also explains the source of the genes that provide the heavy and light chains of the single chain antibody, citing the PhD thesis of Gideon Gross, a co-inventor, which cites a reference providing the complete sequence of the Sp6 light chain gene used to construct the single-chain antibody. Eshhar states that the structure of the Sp6 heavy chain antibody was well known to those of skill in the art and readily accessible on the internet in a database as entry EMBL:MMSP6718. Example 5 at page 54 of the Eshhar specification cites Ravetch et al., J. Exp. Med., 170:481-497 (1989) for the method of producing the CD16α DNA clone that was PCR amplified; this reference published the complete DNA sequence of the CD16α chain, as discussed in paragraph 43 of the Eshhar Declaration. Example 3 of the Eshhar specification uses the DNA of the monoclonal anti-HER2 antibody and states that the N29 hybridoma that produces this antibody was deposited with the Collection Nationale de Cultures de Microorganismes, Institut Pasteur, Paris, on August 19, 1992, under Deposit No. CNCM I-1262. It is incorrect to criticize the methods, examples, and referenced prior art of the Eshhar specification as but "a few PCR primers and probes," as does the Director's brief.

Capon's Example 3 provides a detailed description of the creation and expression of single chain antibody fused with T-cell receptor zeta chain, referring to published vectors and procedures. Capon, like Eshhar, describes gene segments and their ligation to form chimeric genes. Although Capon includes fewer specific examples in his specification than does Eshhar, both parties used standard systems of description and identification, as well as known procedures for selecting, isolating, and linking known DNA segments. Indeed, the Board's repeated observation that the full scope of all of the claims appears to be "enabled" cannot be reconciled with the Board's objection that only a "general plan" to

combine unidentified DNA is presented. See In re Wands, 858 F.2d 731, 736-37 (Fed. Cir. 1988) (experimentation to practice invention must not be "undue" for invention to be considered enabled).

The PTO points out that for biochemical processes relating to gene modification, protein expression, and immune response, success is not assured. However, generic inventions are not thereby invalid. Precedent distinguishes among generic inventions that are adequately supported, those that are merely a "wish" or "plan," the words of Fiers v. Revel, 984 F.2d at 1171, and those in between, as illustrated by Noelle v. Lederman, 355 F.3d at 1350; the facts of the specific case must be evaluated. The Board did not discuss the generic concept that both Capon and Eshhar described -- the concept of selecting and combining a gene sequence encoding the variable domain of an antibody and a sequence encoding a lymphocyte activation protein, into a single DNA sequence which, upon expression, allows for immune responses that do not occur in nature. The record does not show this concept to be in the prior art, and includes experimental verification as well as potential variability in the concept.

Whether the inventors demonstrated sufficient generality to support the scope of some or all of their claims, must be determined claim by claim. The Board did not discuss the evidence with respect to the generality of the invention and the significance of the specific examples, instead simply rejecting all the claims for lack of a complete chimeric DNA sequence. As we have discussed, that reasoning is inapt for this case. The Board's position that the patents at issue were merely an "invitation to experiment" did not distinguish among the parties' broad and narrow claims, and further concerns enablement more than written description. See Adang v. Fischhoff, 286 F.3d 1346, 1355 (Fed. Cir.

2002) (enablement involves assessment of whether one of skill in the art could make and use the invention without undue experimentation); In re Wright, 999 F.2d 1557, 1561 (Fed. Cir. 1993) (same). Although the legal criteria of enablement and written description are related and are often met by the same disclosure, they serve discrete legal requirements.

The predictability or unpredictability of the science is relevant to deciding how much experimental support is required to adequately describe the scope of an invention. Our predecessor court summarized in In re Storrs, 245 F.2d 474, 478 (CCPA 1957) that "[i]t must be borne in mind that, while it is necessary that an applicant for a patent give to the public a complete and adequate disclosure in return for the patent grant, the certainty required of the disclosure is not greater than that which is reasonable, having due regard to the subject matter involved." This aspect may warrant exploration on remand.

In summary, the Board erred in ruling that §112 imposes a *per se* rule requiring recitation in the specification of the nucleotide sequence of claimed DNA, when that sequence is already known in the field. However, the Board did not explore the support for each of the claims of both parties, in view of the specific examples and general teachings in the specifications and the known science, with application of precedent guiding review of the scope of claims.

We remand for appropriate further proceedings.

VACATED AND REMANDED

EXHIBIT
C

JOURNAL OF BACTERIOLOGY, Oct. 1997, p. 6228–6237
0021-9193/97/$04.00+0

Vol. 179, No. 20

Methods for Generating Precise Deletions and Insertions in the Genome of Wild-Type *Escherichia coli*: Application to Open Reading Frame Characterization

ANDREW J. LINK,[1]† DERETH PHILLIPS,[1] AND GEORGE M. CHURCH[1,2]*

Department of Genetics[1] and Howard Hughes Medical Institute,[2] Harvard Medical School, Boston, Massachusetts 02115

Received 6 March 1997/Accepted 8 August 1997

We have developed a new system of chromosomal mutagenesis in order to study the functions of uncharacterized open reading frames (ORFs) in wild-type *Escherichia coli*. Because of the operon structure of this organism, traditional methods such as insertional mutagenesis run the risk of introducing polar effects on downstream genes or creating secondary mutations elsewhere in the genome. Our system uses crossover PCR to create in-frame, tagged deletions in chromosomal DNA. These deletions are placed in the *E. coli* chromosome by using plasmid pKO3, a gene replacement vector that contains a temperature-sensitive origin of replication and markers for positive and negative selection for chromosomal integration and excision. Using kanamycin resistance (Kn^r) insertional alleles of the essential genes *pepM* and *rpsB* cloned into the replacement vector, we calibrated the system for the expected results when essential genes are deleted. Two poorly understood genes, *hdeA* and *yjbJ*, encoding highly abundant proteins were selected as targets for this approach. When the system was used to replace chromosomal *hdeA* with insertional alleles, we observed vastly different results that were dependent on the exact nature of the insertions. When a Kn^r gene was inserted into *hdeA* at two different locations and orientations, both essential and nonessential phenotypes were seen. Using PCR-generated deletions, we were able to make in-frame deletion strains of both *hdeA* and *yjbJ*. The two genes proved to be nonessential in both rich and glucose-minimal media. In competition experiments using isogenic strains, the strain with the insertional allele of *yjbJ* showed growth rates different from those of the strain with the deletion allele of *yjbJ*. These results illustrate that in-frame, unmarked deletions are among the most reliable types of mutations available for wild-type *E. coli*. Because these strains are isogenic with the exception of their deleted ORFs, they may be used in competition with one another to reveal phenotypes not apparent when cultured singly.

With the completion of the *Escherichia coli* K-12 genome sequence (http://www.genetics.wisc.edu/ and http://mol.genes.nig.ac.jp/ccoli/), a variety of tools will be required to determine the functions of the vast array of uncharacterized open reading frames (ORFs) found within the genome. Even in an organism as well studied as *E. coli*, over 58% of the putative coding regions remain without a recognized function, and many others are only partially understood. To study these regions, we devised a system for creating in-frame deletions of any desired sequence in wild-type *E. coli*.

Gene replacements in *E. coli* have generally relied on specific genetic backgrounds as starting strains, such as *polA*, *recD*, *strR*, *sup*$^+$, or F′ (15, 21, 37, 39, 44). After replacement of a wild-type sequence with an in vitro-altered sequence in a mutant background, the altered chromosomal region must then be transduced into a wild-type genetic background. Unfortunately, these methods often require the transduction of a marker along with the mutant allele. This marker can obscure the phenotype of the mutant allele because it may itself cause a mutant phenotype.

Bacterial genes needed in a particular pathway tend to be grouped in cotranscribed clusters or operons (32). Insertional, frameshift, nonsense, or antisense disruption of an ORF within an operon can affect upstream and downstream gene expression in addition to the gene targeted for inactivation. These polar effects could confuse the assignment of a mutant phenotype to the disrupted gene. At the other extreme, point mutants can leave significant parts of the gene intact. To reduce these problems, we developed methods for creating precisely engineered deletions of *E. coli* ORFs by using a procedure known as crossover PCR (18, 19). To integrate these PCR-generated deletions into the genome of wild-type *E. coli*, we constructed a new gene replacement vector, pKO3.

Hamilton et al. have described a method for gene replacement in wild-type *E. coli* that uses homologous recombination between the bacterial chromosome and a plasmid carrying cloned chromosomal sequences whose replication ability is temperature sensitive (16). At the nonpermissive temperature, cells maintain drug resistance only if the plasmid integrates into the chromosome by homologous recombination between the cloned fragment and the bacterial chromosome. Excision of the integrated plasmid is allowed at the permissive temperature. Depending on the position of the second recombination event that excises the plasmid, the chromosome retains either the wild-type sequence or the altered sequence from the plasmid. Although this method can be applied to wild-type strains, there is no selection for loss of the excised plasmid. The *Bacillus subtilis* gene *sacB* encodes levansucrase, an enzyme that catalyzes the hydrolysis of sucrose and levan elongation (12). When expressed in *E. coli* growing on media supplemented

* Corresponding author. Mailing address: Department of Genetics, Warren Alpert Building, Room 513, Harvard Medical School, 200 Longwood Ave., Boston, MA 02115. Phone: (617) 432-7562. Fax: (617) 432-7266. E-mail: church@salt2.med.harvard.edu.

† Present address: Department of Molecular Biotechnology, University of Washington, Seattle, WA 98195.

with sucrose, the *sacB* gene is lethal (14). Blomfield et al. developed a system for using a temperature-sensitive plasmid and a counterselectable *sacB* marker in the chromosome to facilitate allelic exchange (5). We have reduced this system to one component by incorporating the *sacB* gene into a gene replacement plasmid (pKO3) and have developed a protocol for introducing altered alleles into wild-type *E. coli* strains. By combining the crossover PCR and gene replacement methods, we demonstrate a system for creating precise deletions that eliminate gene function without introducing polar affects on expression of distal genes in an operon.

When making a survey of the most abundant proteins in *E. coli*, we found two poorly understood genes, *yjbJ* and *hdeA*, that encode unexpectedly abundant proteins in the cell (25). The *E. coli yjbJ* gene, with sequence similarity to the uncharacterized ORF *ywmH* in *B. subtilis*, encodes a small 69-amino-acid protein that is highly abundant during early stationary phase in rich media. HdeA is a 121-amino-acid protein with a 23-amino-acid signal peptide whose expression is affected by mutations that eliminate the protein HU-1 (45, 46). The HdeA protein is abundant during growth in minimal media and during stationary phase in rich media (25). To determine if mutant alleles of *yjbJ* and *hdeA* have significant phenotypes, we replaced the chromosomal genes with both insertion and deletion alleles by using the pKO3 gene replacement protocol. In the following text, we will discuss the advantages and disadvantages of both insertional and deletion methods. In light of the completion of the genomic sequences of several free-living organisms, the results of these gene replacements are discussed as paradigms for addressing the function of chromosomal sequences.

MATERIALS AND METHODS

Strains. All plasmid constructions were electroporated and propagated in *E. coli* DH5α [F⁻ λ⁻ *endA1 hsdR17 hsdM⁺ supE44 thi1 recA1 gyrA96 relA1* Δ(*argF lacZYA*)U169 φ80d Δ(*lacZ*)M15]. The gene replacement experiments used the recombination proficient wild-type K-12 strain EMG2 (F⁺ λ⁺).

Media and growth conditions. All strains were grown in LB medium (1% Bacto Tryptone, 0.5% yeast extract, 0.5% NaCl) with the appropriate selection. For antibiotic selection, the concentrations of antibiotics were 50 mg/ml (ampicillin and kanamycin) and 20 mg/ml (chloramphenicol). For selection against *sacB*, LB medium was supplemented with sucrose to a final sucrose concentration of 5% (wt/vol).

DNA purification. Plasmid DNA was purified by the alkaline lysis method (4). Genomic DNA was purified by previously described methods (10).

Partial digestion. All partial digestion of genomic and plasmid DNA used serial dilution of the restriction enzyme and a constant 1-h incubation time (26). The reactions were stopped by adding 0.25 M EDTA (pH 8) to a final concentration of 50 mM.

Blunt-end reactions. Unless stated otherwise, T4 DNA polymerase and deoxynucleoside triphosphates (dNTPs) were used to create all blunt-ended DNA fragments.

Ligation. Ligations were performed overnight at room temperature, using a DNA concentration of 10 μg/ml and an insert-to-vector molar ratio of 1:1 or an oligonucleotide-to-insert molar ratio of 160:1. The ligation buffer used for the reactions contained 66 mM Tris-HCl (pH 7.5), 5 mM $MgCl_2$, 50 mM dithiothreitol, 1 mM ATP, and 0.05 Weiss units of T4 DNA ligase/μl. The ligated DNA was ethanol precipitated, washed with 70% ethanol, vacuum dried, and resuspended either in 10 mM Tris-HCl (pH 8)–1 mM EDTA or 40 μl of electroporation-competent *E. coli* cells (for immediate transformation).

Electroporation. Electroporation-competent cells (40 μl; 10^{11} CFU/ml) were mixed with 1 to 3 μl of DNA solution in an ice-cold microcentrifuge tube and transferred to a 0.2-cm electroporation cuvette (Bio-Rad, Inc.). The cells were electroporated at 2.5 kV with 25 μF and resistance of 200-ohms. Immediately after electroporation, 1 ml of SOC medium (2% Bacto Tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, 20 mM glucose) was added to the cuvette. The cells were transferred to a 17- by 100-mm polypropylene tube and allowed to recover for 1 h at either 30°C (for temperature-sensitive plasmids) or 37°C with shaking at 250 rpm before plating on selective media.

PCR. All PCRs were performed in a Perkin-Elmer 9600 thermal cycler. PCR buffer (28) consisted of 30 mM tricine (pH 8.4), 2 mM $MgCl_2$, 5 mM β-mercaptoethanol, 0.01% (wt/vol) gelatin, 0.1% (wt/vol) Thesit, 200 μM each dNTP, 600 μM each primer, and 1 U of *Taq* polymerase (Boehringer Mannheim, Inc.). After addition of template DNA, the PCR mixture was denatured at 94°C for 3 min before addition of the *Taq* polymerase. The thermal cycle profile was 15 s at 94°C, 15 s at 55°C, and 30 s at 72°C. All experiments used 30 cycles and a final 5-min 72°C hold step.

Analysis of PCR products. PCR products were analyzed on 2% high-strength agarose–1% NuSieve agarose gels (FMC, Inc.) or 1% high-strength agarose gels cast in 0.5× Tris-borate-EDTA with ethidium bromide.

pKO3 plasmid construction. The gene replacement vector pKO3 was constructed as follows. First, the 1.6-kb *Eco*NI fragment from pMAK700 (16) containing the temperature-sensitive pSC101 replication origin and the 1.6-kb *Hin*dIII-*Bsu*36I fragment of pMAK700 containing the *cat* gene were blunt ended and ligated together to create pKO1.

Second, the 1.35-kb *Not*I-*Nru*I fragment from pBS-TS (2a) containing the *sacB* gene and the 5.6-kb *Sph*I-linearized pMAK705 plasmid (16) were blunt ended and ligated to create pMAK705s. The following *Not*I polylinker was then ligated into the *Bam*HI site of pMAK705s to create pMAK705so:

5′-GATCGCGGCCGCGGGACCCCATCCTCTAGAGCGGCCGC-3′
3′-CGCCGGCCGCCTGGGCCTAGGAGATCTCGCCGGCGCTAG-5′

The 550-bp *Bgl*I-*Bsm*AI fragment from pBluescript II SK− (Stratagene, Inc.) containing the M13 origin of replication and *Hin*dIII-linearized plasmid pMAK705so were blunt ended and ligated to create pMAK705som. The single *Aat*II site in plasmid pMAK705som was deleted by using T4 DNA polymerase and dNTPs to create pMAK705somp.

Finally, the 2.4-kb *Eci*136II-*Eco*RV fragment from pMAK705somp containing the polylinker, M13 origin of replication, and *sacB* gene was blunt ended and ligated to *Age*I-linearized, blunt-ended plasmid pKO1 to create pKO3.

Crossover PCR deletions and subcloning. Crossover PCR deletion products were constructed in two steps, as illustrated in Fig. 4. In the first step, two different 25-μl asymmetric PCRs were used to generate fragments to the left and right of the sequences targeted for deletion. The PCR conditions were as described above except that the primer pairs were in a 10:1 molar ratio (600 μM outer primer and 60 μM inner primer). In the second step, the left and right fragments were annealed at their overlapping region and amplified by PCR as a single fragment, using the outer primers. Specifically, 1 μl of each of the two asymmetric PCR mixtures and 600 μM each of the two outside primers were mixed together and PCR amplified. The fusion products were phenol-chloroform extracted, ethanol precipitated, washed with 70% ethanol, vacuum dried, resuspended in 50 μl of 1× *Bam*HI restriction buffer containing 40 U of *Bam*HI restriction enzyme, and digested overnight at 37°C. The fusion products were gel purified, ligated into *Bam*HI-digested and phosphatase-treated pKO3 vector, electroporated into *E. coli*, and plated on chloramphenicol plates at 30°C. The recombinant colonies were screened for inserts with PCR, using primers pKO3-L and pKO3-R (described below).

To construct the 286-bp deletion of *yjbJ* by crossover PCR, the following set of oligonucleotide primers was used: yjbJ-No, 5′-CGCGGATCCTCACCTTTAC CGCCTATGCG-3′; yjbJ-Ni, 5′-CCCATCCACTAAACTTAAACACCGTCA CGTTGCGGCAAACC-3′; yjbJ-Co, 5′-CGCGGATCCTTGCGCCTGATGAG TCTGCAGG-3′; and yjbJ-Ci, 5′-TGTTTAAGTTTAGTGGATGGGGTGGAT TGGGAAACCCGC-3′.

To construct the deletion of *hdeA*, the following set of primers was used: hdeA-No, 5′-CGCGGATCCGAAATTATGACTGCGGTTGC-3′; hdeA-Ni, 5′-CCCATCCACTAAACTTAAACAGCCTAATACTTTTTTCATCG-3′; hdeA-Co, 5′-CGCGGATCCTACTCCTTTTTACTTGCACC-3′; and hdeA-Ci, 5′-TG TTTAAGTTTAGTGGATGGGAAAGGCGAATGGGACAAAAT-3′.

DNA sequencing. DNA sequencing was performed as previously described with the Stratagene Cyclist sequencing kit II (27). Sequencing products were labeled with [α-^{32}P]dATP and resolved on a 4.5% wedge-gradient sequencing gel. Sequencing primers used for the pKO3 left and right vector-insert junctions were pKO3-L (5′-AGGGCAGGGTCGTTAAATAGC-3′) and pKO3-R (5′-T TAATGCGCCGCTACAGGGCG-3′). Sequencing primers used to prime from multiplex tag 04 (10) were CP-04 (5′-AGTGTGAGGTTTAAATATTG-3′) and CE-04 (5′-TGTTTAAGTTTAGTGGATGG-3′). Sequencing primers used to prime from multiplex tag 01 (10) were CP-01 (5′-TGATTAGTTGTAATGAA AGG-3′) and CE-01 (5′-TAGTATGATTTTATTGGGGG-3′).

Gene replacement. Mutant alleles cloned into the pKO3 gene replacement vector were electroporated into EMG2 and allowed to recover for 1 h at 30°C. The cells were plated on prewarmed chloramphenicol-LB plates and incubated at 43 and 30°C. The integration frequency was calculated as the ratio of colonies at 43°C to colonies at 30°C. From the 43°C plate, one to five colonies were picked into 1 ml of LB broth, serially diluted, and immediately plated at 30°C on either 5% (wt/vol) sucrose or 5% sucrose-kanamycin plates and at 43°C on chloramphenicol plates. The excision frequency is the ratio of 30°C-grown sucrose-resistant colonies to 43°C-grown chloramphenicol-resistant colonies. The 5% sucrose plates were replica plated to chloramphenicol plates at 30°C to test for loss of the replacement vector. The gene replacement was confirmed by PCR using primers flanking the targeted ORF.

Construction of multiplex interposons. To construct the kanamycin resistance (Kn^r) interposon, the 1.3-kb *Dra*III-*Bam*HI fragment from pNK2859 (23) containing the *kan* gene was blunt ended and ligated to the following *Bst*XI linkers:

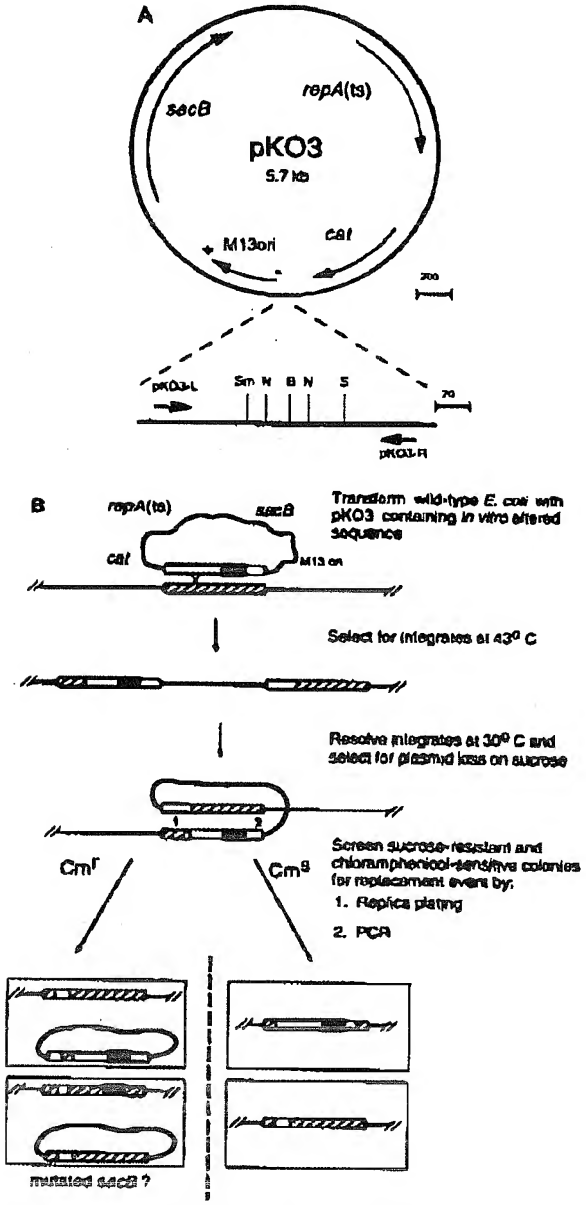

FIG. 1. Gene replacement vector and protocol. (A) The pKO3 vector used in the gene replacement experiments. The cloning region is enlarged. Arrows in the circular plasmid indicate the direction of transcription and the direction of M13 replication. The arrows in the enlarged region are the DNA primer sites. Unique restriction sites are shown (B, *Bam*HI; N, *Not*I; S, *Sal*I; Sm, *Sma*I). ori, origin of replication. (B) Protocol used for replacing wild-type sequences on the chromosome with in vitro-altered sequences. The gene replacement vector carrying in vitro-altered sequences is transformed into *E. coli* and plated at the nonpermissive temperature of the plasmid replicon. An integration event allows replication of plasmid sequences by the chromosomal origin. When shifted to 30°C, the plasmid is excised from the chromosome at either crossover point 1 or 2. The counterselectable *sacB* marker is used to select for loss of plasmid sequences. The sucrose-resistant colonies are screened for loss of vector sequences by replica plating to chloramphenicol plates and then for the gene replacement event by PCR. The "mutated *sacB*?" in the left panels indicates loss of *sacB* gene function by some unknown mechanism. The wavy, thin line represents the gene replacement vector sequences. The straight, thin line represents the *E. coli* chromosome. The boxes represent homologous sequences cloned into the vector (open) and located in the *E. coli* chromosome (striped). The black box within the homologous vector sequence could represent any type of sequence alteration (insertion, deletion, single-base change, etc.).

5'-TCTAGACCACCTGC-3'
3'-AGATCTGGTG-5'

The *kan* fragment with the attached *Bst*XI linkers was ligated to the 2.5-kb *Bst*XI fragment from multiplex vector plex.04B containing multiplex tags 04 (10) to create plasmid pplexkan04B. The Kn^r fragment with the attached linkers was similarly inserted into multiplex vectors 01, 02, 07, 09, 10, 11, 14, 16, 17, 18, and 19 (10) to construct a series of plexkan interposons.

To add a 5'-CG-3' overhang on the Kn^r interposon, the 1.4-kb *Not*I fragment from plasmid pplexkan.04B was ligated with the following adapters and gel purified:

Dap1 5'-CGCCCCCTGCAGGA-3'
Dap2 3'-GGGGGACGTCCTCCGG-5'

Constructing *pepM*, *rpsB*, *yjbJ*, and *hdeA* insertion mutations. A gel-fractionated genomic library of 3- to 7-kb *Sau*3A inserts prepared from EMG2 genomic DNA was ligated into the phosphatase-treated *Bam*HI site of pKO3. This library was electroporated into *E. coli* DH5α, plated on chloramphenicol plates, and overlaid with nylon membranes to create colony lifts essentially as previously described (26). To identify recombinant plasmids carrying the desired genomic inserts, the colonies were screened by hybridization using oligonucleotides labeled at the 5' end with [γ-^{32}P]ATP and T4 polynucleotide kinase at a probe concentration of 1 nM, overnight at 42°C (10, 26). The probes used were complementary to the 5' ends of *pepM* (5'-TTCTGTCCATCAGCGTCGGTG-3'), *yjbJ* (5'-GCCGGCTTCATCTTTATTCAT-3'), and *hdeA* (5'-CCACCAAGAATAACGCCTAAT-3'). The *pepM* oligonucleotide was used to screen for *rpsB* clones simultaneously.

To identify clones with at least 1 kb of genomic DNA flanking each side of the desired genes, positive clones were screened by a combination of restriction mapping and DNA sequencing across the vector-insert junctions. These results were compared to physical maps of the regions. To create lesions in *pepM* and *rpsB*, a positive clone was partially digested with a mixture of five four-base recognition site restriction enzymes that create a 5'-CG-3' overhang (*Acl*I, *Hpa*II, *Hin*pI, *Mac*II, and *Taq*I). The singly cut, linearized plasmid was gel purified and ligated with the Kn^r interposon plexkan04, using adapters Dap1 and Dap2. For *yjbJ* and *hdeA*, positive clones (p1.7 [*yjbJ*] and p15.3 [*hdeA*]) were linearized by partial digestion with restriction enzymes with unique sites in the ORFs (e.g., *Noe*I for *yjbJ* and *Pst*I or *Pvu*II for *hdeA*). The singly cut, linearized plasmids were gel purified and ligated to a blunt-ended multiplex interposon (e.g., *yjbJ*::plexkan04, *hdeA*::plexkan01 [*Pst*I site], and *hdeA*::plexkan04 [*Pvu*II site]). Before performing the gene replacement, we characterized the *in vitro*-altered insert by DNA sequencing across both the vector-insert junctions (using the primer sites in the vector as primer sites) and the interposon-insert junctions (using the interposon's multiplex tags as primer sites).

Screening for gene replacements. PCR was used to screen for gene replacements of *yjbJ* and *hdeA*. The *yjbJ* gene replacement was confirmed by using the primers yjbJ-Nout and yjbJ-Cout flanking the gene. The *hdeA* gene replacement was confirmed by PCR using primer pair hdeA-Nout1 plus hdeA-Cout1, hdeA-Nout2 plus hdeA-Cout2, or hdeA-Nout3 plus hdeA-Cout3 flanking the gene. Sequences of the primers are as follows: yjbJ-Nout, 5'-AGGTGAAAAAGAAACCGCGTT-3'; yjbJ-Cout, 5'-TGGTTTGCCGCAACGTGACGG-3'; hdeA-Nout1, 5'-CGCGGATCCCATATACAGAAAACC-3'; hdeA-Cout1, 5'-CGCGGATCCTTTTAAAGAAGATAT-3'; hdeA-Nout2, 5'-CTGATGCATCTGTAACTCATT-3'; hdeA-Cout2, 5'-AACGCAGATTGTGCGTTCACC-3'; hdeA-Nout3, 5'-GGATGAAGAAATAGCCGATC-3'; and hdeA-Cout3, 5'-CTTCCCATGCCAATTAATAC-3'.

Competition experiments. Competition experiments were performed by coculturing equal concentrations of two strains in rich media and then sampling the population density of each strain at various time points. Equal optical densities at 600 nm of diluted overnight cultures of the various strains were mixed in the following combinations and sampled at various time points. EMG2 *yjbJ*::plexkan04 and EMG2 *hdeA*::plexkan01 strains were each cocultured with the wild-type EMG2 strain. In a second competition experiment, EMG2 *yjbJ*::plexkan04 was cocultured with EMG2 Δ*yjbJ*. Each mixed culture was grown aerobically in a 250-ml Erlenmeyer flask containing 50 ml of LB medium at 37°C shaking at 250 rpm (New Brunswick Scientific G2 platform). Since each culture contained both a marked and unmarked strain, the survival ratios could be determined by plating on both LB and kanamycin plates at various time points and counting the colonies surviving on each plate.

RESULTS

Developing an improved gene replacement method. We constructed a gene replacement vector for creating null mutations in the chromosomal sequences of wild-type *E. coli* strains as described in Materials and Methods and illustrated in Fig. 1A. The plasmid is derived from a previously described gene replacement vector and has the *lac* sequence removed to eliminate homologous recombination at the *lac* region in the *E. coli* chromosome (16). The *repA*(Ts) replication origin is derived from pSC101 and has a permissive temperature of 30°C but is inactive at 42 to 44°C. The *cat* gene (encoding chloramphenicol resistance) is used as a marker to select for chromosomal integrates and as a marker for cells harboring vector sequences after plasmid excision. The *sacB* gene is used to counterselect vector sequences by growing cells harboring the plasmid on medium supplemented with 5% sucrose. The M13 replication origin facilitates generation of single-stranded copies of the plasmid by using helper phage (a feature not used in this study). Finally, the primer sites pKO3-L and pKO3-R flanking the cloning site enable screening the vector for inserts by PCR or for DNA sequencing across the vector-insert junctions.

Figure 1B diagrams the protocol that we used to perform gene replacements in *E. coli*. The in vitro-altered sequences carried in the vector pKO3 are transformed into *E. coli*, and the transformed cells are allowed to briefly recover at the permissive temperature. The cells are then plated on chloramphenicol plates at the nonpermissive temperature to select for chromosomal integrates. This was more effective for obtaining the final gene replacement event than plating cells at 30°C and shifting them to 43°C. We found that integrates could also be obtained by serially diluting cells harboring the plasmid at 30°C and plating them at 43°C. To select cells in which the plasmids are excised and lost, we picked and suspended colonies from the 43°C plates, diluted the suspension, and plated the cells on LB plates containing 5% sucrose at 30°C. Only cells that have excised the plasmid sequences and lost *sacB*'s counterselectable function should grow under these conditions. We found this procedure worked better for getting the final gene replacement event than simply replica plating colonies from 43°C to sucrose plates at 30°C. Finally, the sucrose-resistant and chloramphenicol-sensitive colonies are screened for the desired gene replacement event by using PCR and primers to the genomic DNA flanking the altered sequences or by Southern hybridization.

Replacing *yjbJ* with an insertional allele. Suspecting that the null allele of *yjbJ* would be lethal, we decided to disrupt *yjbJ* by inserting a specialized Kn^r selectable marker, or interposon, into the gene (29). A 5.5-kb DNA fragment from a genomic library containing *yjbJ* was cloned into pKO3, and a Kn^r gene (plexkan04) was inserted at the unique *Nae*I site in the gene (see Materials and Methods). Before doing the gene replacement, we sequenced both the vector-insert junctions and the insertion site of the Kn^r gene and showed that the insertional allele had at least 1 kb of chromosomal sequence flanking both sides of the interposon (Fig. 2A). When the *yjbJ* replacement vector was transformed into *E. coli* and plated at 43°C, the integration frequency was 10^{-2} of the plated cells. Several integrates were picked, serially diluted, and plated at 30°C on various selective media to induce the plasmid excision and loss (Fig. 2B). These different master plates were then replica plated to chloramphenicol plates and kanamycin plates to identify colonies that retained the Kn^r gene and not the vector (Fig. 2B). When the integrate cells were plated on kanamycin medium without sucrose at 30°C, most of the sucrose-resistant colonies were still chloramphenicol resistant, indicating that the cells retained the vector sequences (Fig. 2B, row a). When integrate cells were plated on kanamycin–5% sucrose medium, more than 98% of the sucrose-resistant colonies were chloramphenicol sensitive, indicating loss of plasmid sequences and a probable gene replacement event (Fig. 2B, row b). When the integrate cells were plated on rich medium containing 5% sucrose, 48% of the sucrose-resistant colonies were chloramphenicol sensitive and kanamycin resistant, indicating loss of the plasmid and a probable gene replacement event (Fig. 2B, row c). We verified the structure of the initial 43°C integration and the replacement of *yjbJ* with the insertional allele by screening colonies via PCR using primers flanking *yjbJ* (Fig. 2C). These results proved that the pKO3 replacement system worked and showed that *yjbJ* is a nonessential gene under these environmental conditions.

Lethal gene replacement phenotype. To observe the results of the gene replacement protocol when trying to replace an essential *E. coli* gene with an insertional allele, we tested two known essential genes, *pepM* and *rpsB*. The *pepM* (*map*) gene encodes methionine aminopeptidase, and *rpsB* encodes the ribosomal protein S2 (7, 9). Each gene was cloned into the gene replacement vector pKO3, and insertion mutations were constructed by using the Kn^r gene (see Materials and Methods). DNA sequencing and restriction enzyme mapping showed that both inserts had at least 1 kb of genomic DNA flanking each side of the insertion site.

The pKO3 plasmids carrying the insertionally disrupted essential genes were electroporated into *E. coli* and plated at 43°C to select for integration. The integration frequency was approximately 10^{-2} to 10^{-3}, similar to that for *yjbJ*. Ten integrate colonies were picked, suspended in medium, serially diluted, and plated at 30°C on 5% sucrose–kanamycin plates. We found the sucrose resistance frequencies for both the *pepM* and *rpsB* integrates were approximately 10^{-5}, compared to a frequency of 10^{-2} to 10^{-4} for the insertional disrupted nonessential *yjbJ* gene replacement. For both *pepM* and *rpsB*, all of the sucrose-resistant, kanamycin-resistant colonies remained chloramphenicol resistant, indicating that plasmid sequences were still present in the cell. In addition, the colonies had a mucoid phenotype compared to colonies that had lost the plasmid sequences. It is unknown whether *sacB*'s activity had been directly compromised by a mutation in the gene or if a secondary mutation in the genome conferred sucrose resistance. These results showed the phenotype expected when one tries to replace an essential gene with a disrupted allele. Using the pKO3 gene replacement procedure, Brown et al. have shown that an essential gene, *murA*, can be replaced on the *E. coli* chromosome with a deletion allele, as long as the deletion is complemented by another copy of the essential gene (8).

Paradoxical phenotypes of different *hdeA* insertional alleles. Speculating that *hdeA* might be an essential gene, we constructed two different insertional alleles of *hdeA* (Fig. 3). Both were made using the same chromosomal insert cloned into the vector pKO3. In one allele, the inserted Kn^r gene was cloned into the *Pst*I site of the *hdeA* gene; in the second allele, the Kn^r gene was cloned into the *Pvu*II site in the opposite orientation (see Materials and Methods).

In the first step of the gene replacement procedure, the two plasmids transformed and integrated at similar frequencies. However, when resolving the integrates, we found that the *Pvu*II insertional allele had a sucrose resistance frequency of $<10^{-8}$, compared to approximately 10^{-3} for the *Pst*I allele. All of the sucrose-resistant, kanamycin-resistant colonies with the *Pvu*II insertional allele were chloramphenicol resistant, indicating that the plasmid sequences were still present. This finding suggested that the *Pvu*II allele is a lethal mutation. How-

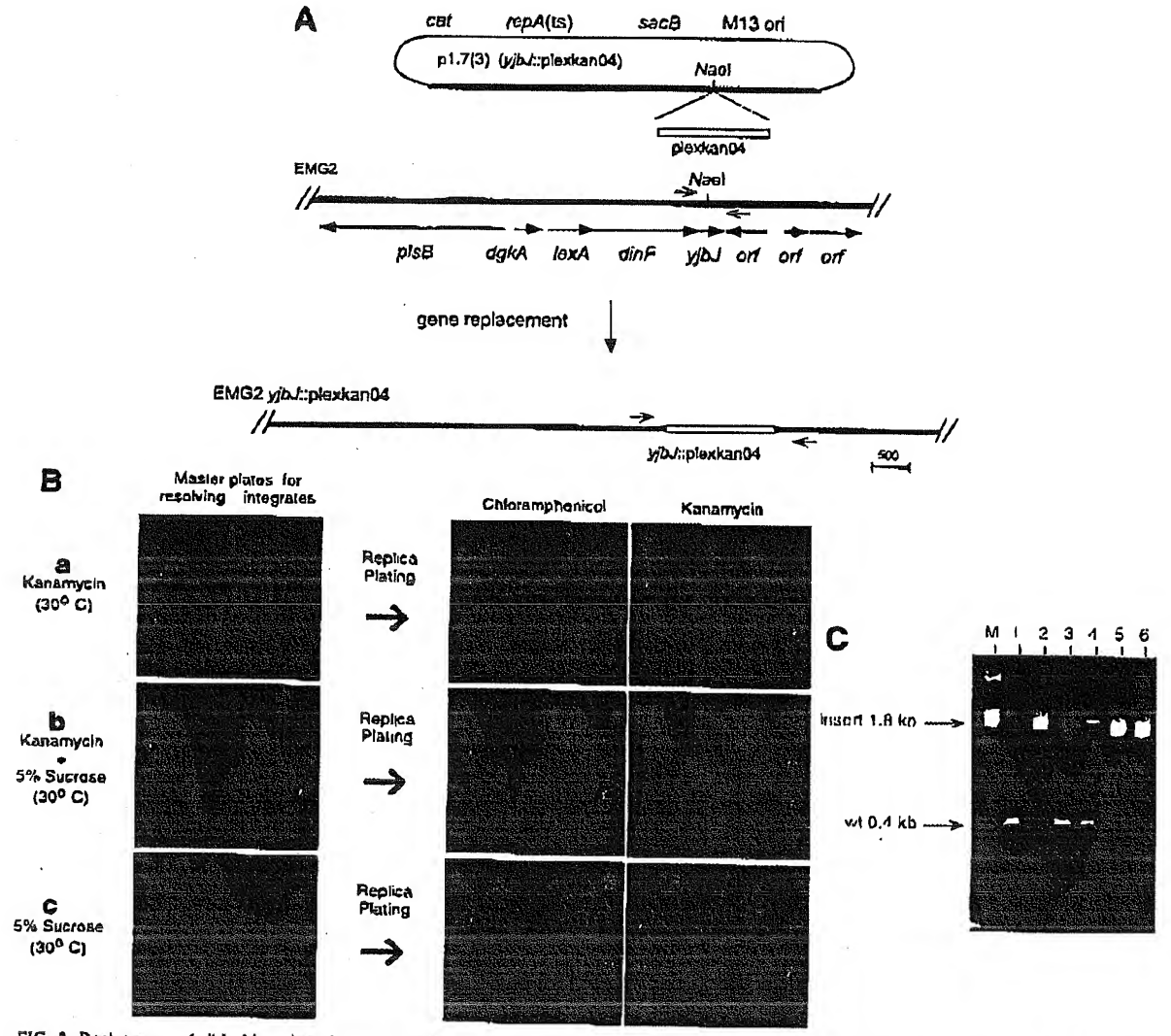

FIG. 2. Replacement of *yjbJ* with an insertional allele. (A) The top panel shows a genomic insert containing *yjbJ* cloned into pKO3 and mutagenized with the plexkan04 interposon. The interposon insertion at the *NaeI* site was confirmed by sequencing across the interposon-genomic insert junction by using primers CE-04 and CP-04, which flank the interposon. The insert was mapped on the *E. coli* chromosome by sequencing across the vector-insert junctions by using primers pKO3-L and pKO3-R. The physical map of the wild-type chromosomal region is shown. The arrows flanking the *NaeI* sites are the PCR primer sites yjbJ-Nout and yjbJ-Cout used to identify the *yjbJ* allele. The expected sizes of the PCR products are 0.4 kb for *yjbJ* and 1.8 kb for *yjbJ*::plexkan04 alleles. The thick lines represent chromosomal fragment sequences. The thin line represents the vector pKO3. The open box represents the interposon. (B) Resolution of *yjbJ*::plexkan04 integrates when plated at 30°C under different selection conditions. Five 43°C integrates were picked, serially diluted, and plated on the master plates shown at the left. These master plates were replica plated first to chloramphenicol plates and then to kanamycin plates at 30°C to detect loss of plasmid sequences (chloramphenicol sensitive) and retention of the interposon (kanamycin resistant). (C) Use of PCR to verify the replacement of *yjbJ* with the insertional allele. The gel shows the products of the PCRs using primers yjbJ-Nout and yjbJ-Cout to detect either the wild-type (wt) or the disrupted *yjbJ* allele. The sources of the template DNA are plasmid p1.7 (*yjbJ*) (lane 1), plasmid p1.7(3) (*yjbJ*::plexkan04) (lane 2), EMG2 genomic DNA (lane 3), genomic DNA from 43°C integrate of p1.7(3) (*yjbJ*::plexkan04) into the EMG2 (lane 4), EMG2 *yjbJ*::plexkan04 genomic DNA from a sucrose-resistant and chloramphenicol-sensitive colony shown in row b (lane 5), and EMG2 *yjbJ*::plexkan04 genomic DNA from a sucrose-resistant and chloramphenicol-sensitive colony shown in row c (lane 6). The size marker is a 123-bp ladder (lane M).

ever, the replacement of the wild-type gene with the *PstI* insertion was confirmed by PCR screening of colonies with primers flanking *hdeA*. The latter finding suggested that the *PstI* allele is nonlethal. These *PstI* and *PvuII* results together illustrated the difficulty of classifying *hdeA* as either an essential or a nonessential gene because the phenotype varied according to the insertion site of the marker and/or its orientation in the disrupted allele.

Engineering in-frame deletions to minimize polar effects. To avoid problems associated with insertional mutations, we developed a system that replaces ORFs with in-frame deletions. Figure 4 shows how we used crossover PCR to create a dele-

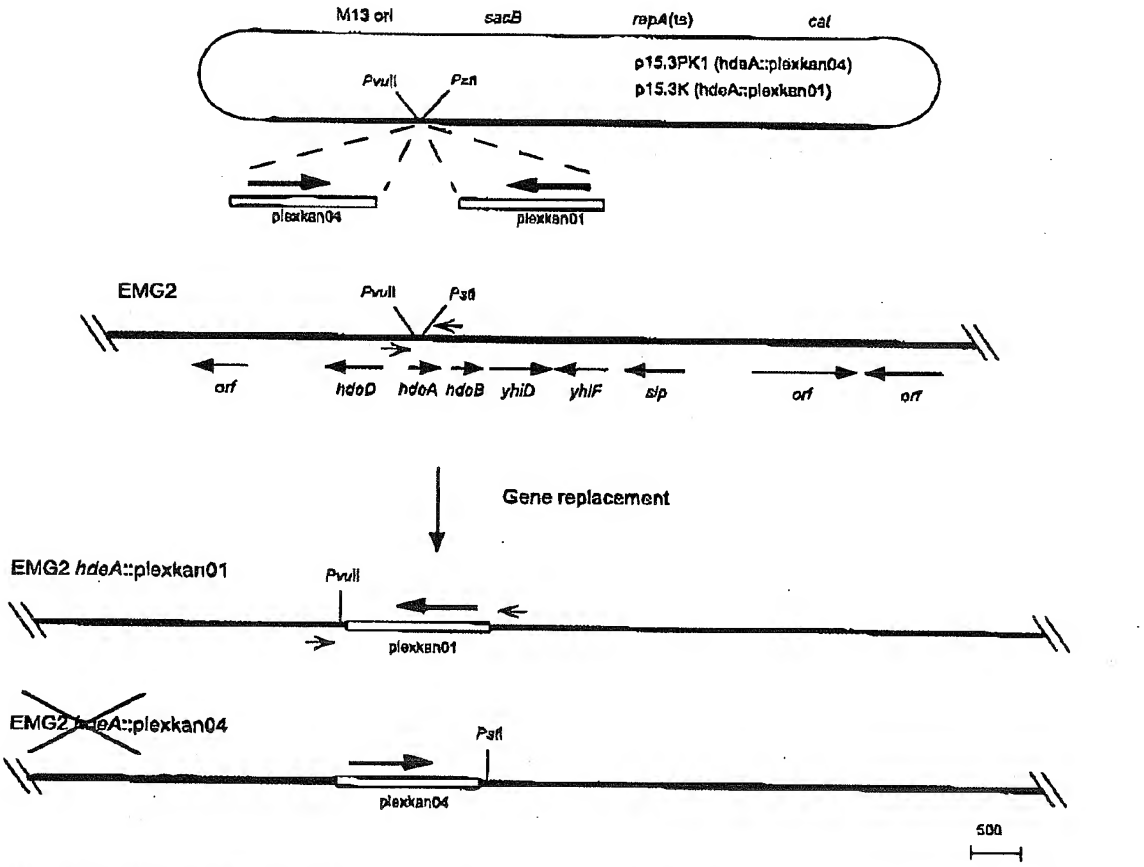

FIG. 3. Paradoxical phenotypes of replacing *hdeA* with insertional alleles. Physical maps of the DNA fragment containing *hdeA* and the chromosomal region before and after gene replacement with the insertional alleles are shown. The unique *PvuII* and *PstI* sites in *hdeA* are the insertion sites for the plexkan01 and plexkan04 interposons, respectively. The arrows flanking the *PstI* and *PvuII* sites are the PCR primer sites hdeA-NotI and hdeA-CoaI flanking the *hdeA* gene. The construct marked with an "X" could be integrated but not resolve to the replacement allele. Details are as described in the legend to Fig. 2A.

tion of any *E. coli* ORF (18, 19). Complementary oligonucleotide primers and asymmetric PCR are used to generate two DNA fragments having overlapping ends. The two fragments are combined in a fusion reaction in which the overlapping ends annealed and served as primers for 3′ extension of the complementary strand. This fusion molecule is then amplified by PCR using the outer primers.

To construct the deletions, we developed the following rules for designing the oligonucleotides to ensure sufficient homology for recombination during gene replacement and to minimize disruption of flanking sequences. The lengths of the two fragments flanking the deletion are at least 500 bp. The decision to use 500 bp is based on published integration frequencies for various lengths of chromosomal regions cloned into similar gene replacement vectors. The predicted integration frequency should be approximately 10^{-4} (5, 16). The two complementary oligonucleotides (C and B) have at least a 21-base complementary region to allow the products from the asymmetric first and second PCRs to anneal and extend (Fig. 4). The primers were designed so that the deletion maintained the original translational reading frame of the ORF and the added bases provided unique sequences for tracking the deletion in a population of different *E. coli* deletion strains. To minimize potential affects on expression of neighboring genes, we engineered the deletion of the ORF to begin 18 bp downstream of the translation start site and end 36 bp upstream of the stop codon. The oligonucleotides (A and D) have *Bam*HI restriction sites in the 5′ end to allow efficient cloning of the fusion product.

Deletion of *hdeA*. To further investigate the null phenotype of the *hdeA* mutation, we engineered a deletion of the gene (Fig. 5A). Using the above-specified rules, we deleted a 279-bp region, or 84% of the coding region of *hdeA* and replaced it with a 21-bp in-frame sequence tag, using the crossover PCR protocol (see Materials and Methods). Figure 5B shows the two complementary PCR products and the final crossover PCR deletion product. The deletion fragment was cloned into the vector pKO3, and chromosomal deletions were introduced into the chromosome by using our pKO3 gene replacement protocol. The deletion plasmid had an integration frequency of 1.8×10^{-3}. Integrates were serially diluted and plated at 30°C on 5% sucrose plates to select for excision and loss of the

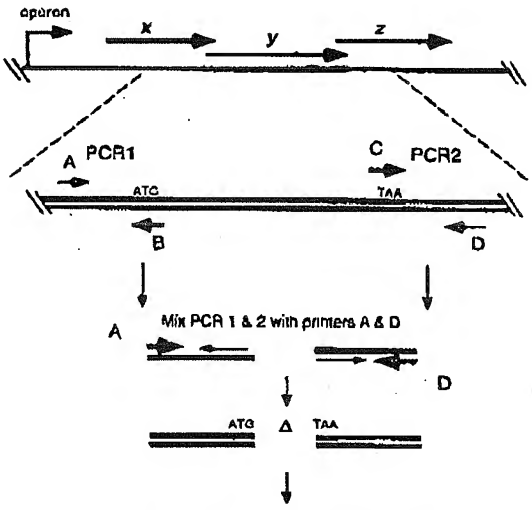

FIG. 4. The creation of in-frame deletion constructs. The top line represents a region of the chromosome where genes *x*, *y*, and *z* form a polycistronic operon. The second line is an expanded view of gene *y* showing the two PCRs used to generate fragments (PCR1 and PCR2) which will form an in-frame deletion of gene *y* when fused. The PCR primers B and C are complementary over 21 nucleotides (represented by the light gray lines) so that when the two PCR products are mixed, the complementary regions anneal and prime at the 3′ overlapping region for a 3′ extension of the complementary strand. In the third line, the fused molecule is amplified by PCR with primers A and D. Primers A and D have *Bam*HI sites incorporated into the 5′ ends of both oligonucleotides (represented by the gray lines) so that the fusion product can be restriction digested and cloned into pKO3.

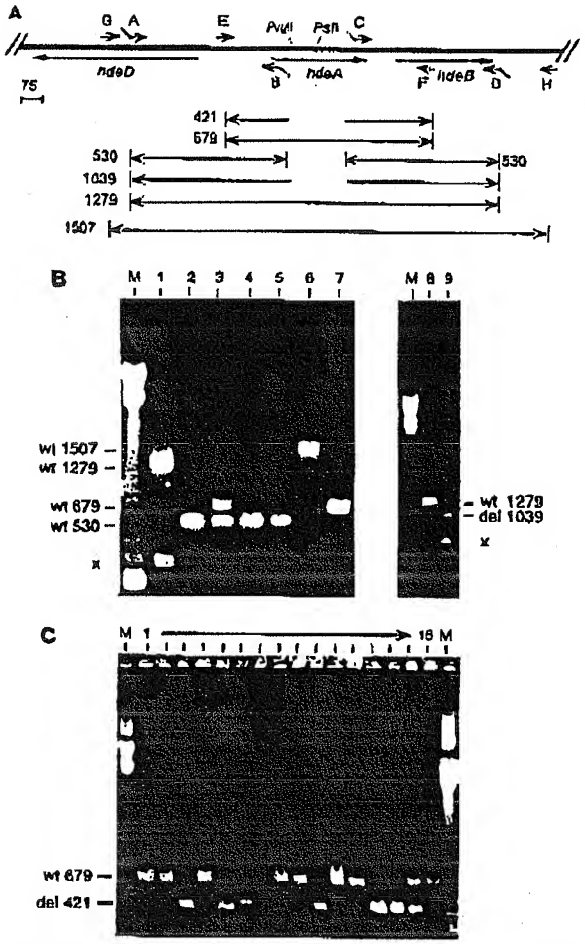

FIG. 5. Constructing and replacing the gene *hdeA* with a precisely engineered deletion. (A) Diagram of the *hdeA* region. The small arrows marked with capital letters are the PCR primer sites used to construct the deletion and to PCR assay either the wild-type or deletion allele of *hdeA*. The predicted sizes of the PCR products are shown below the physical map. Primers: A, hdeA-No; B, hdeA-Ni; C, hdeA-Ci; D, hdeA-Co; E, hdeA-Nout2; F, hdeA-Cout2; G, hdeA-Nout3; H, hdeA-Cout3 (see Materials and Methods). (B) Analysis of PCR products used to construct precise deletion of *hdeA*. The left gel shows the two fragments that will form the deletion product and the PCR products made by using primers flanking *hdeA*. The sizes of the PCR products are shown. EMG2 genomic DNA was used as the template DNA for lanes 1 to 7. The DNA primer pairs used for the PCRs are A-D (lane 1), A-B (1:1) (lane 2), A-B (10:1) (lane 3), D-C (1:1) (lane 4), D-C (10:1) (lane 5), G-H (lane 6), and E-F (lane 7) (see panel A). The right gel shows the amplified deletion (Δ) product (lane 9). Using the products of lanes 3 and 5 as templates, the left and right fragments of the deletion were combined, annealed, extended, and PCR amplified by using primers A and D (lane 9). This gel also shows the PCR product made by using the same primer pairs but starting with EMG2 genomic DNA as the template (lane 8). The deletion fusion product was cloned into pKO3 and used in the gene replacement protocol. The "x" indicates an unknown PCR by-product. The size marker is a 123-bp ladder (lane M). wt, wild type. (C) Verification of the replacement of *hdeA* with the crossover PCR deletion product. After integration at 43°C, integrates were plated at 30°C on 5% sucrose plates and replica plated to chloramphenicol plates. The chloramphenicol-sensitive, sucrose-resistant colonies were screened by PCR using primers E and F (see panel A). Those containing the precise deletion give a 421-bp product, while those containing the wild-type allele give a 679-bp product. This gel shows a subset of the colonies screened for the deletion (lanes 1 to 16). The size marker is a 123-bp ladder (lane M).

plasmid sequence. Figure 5C shows the results of screening a fraction of the resolved colonies by PCR with primers flanking *hdeA*. Approximately 7% of the sucrose-resistant and chloramphenicol-sensitive resolved integrate colonies had the deletion replacing the wild-type *hdeA* sequence. Why the resolution frequency for replacing the gene with the deletion was not the expected 50% is unknown. Recovery of the deletion demonstrated that the gene was nonessential under these environmental conditions and suggested that the apparent lethal effect of the Km^r insertion into *hdeA* was probably due to an effect of the insertion on elements outside of the cloned segment.

Deletion of *yjbJ*. A similar set of experiments was performed to delete 146 bp, or 73%, of *yjbJ*, with the rules previously described. Although the deletion product was successfully amplified, it could not be cloned into pKO3. A PCR assay showed the deletion insert ligated to pKO3, and so we hypothesize that either the protein produced by the deletion mutant was toxic or the insert interfered with plasmid replication in *E. coli*. An analysis of the genomic region identified a potential promoter 107 bp upstream of the *yjbJ* translational start. We designed a second deletion to remove most of the predicted promoter region while leaving the upstream *dinF* gene intact. The second deletion extended from 3 bp downstream of the *dinF* stop codon to 36 bp upstream of the *yjbJ* stop codon, deleting 286 bp of the region, including 82% of coding region of *yjbJ*. This second crossover PCR deletion product was successfully cloned into pKO3. Using the gene replacement protocol, we found that the deletion plasmid had an integration frequency of 7.9×10^{-5}. Similar to *hdeA*, only 3% of the sucrose-resistant

and chloramphenicol-sensitive resolved-integrate colonies had the deletion replacing the wild-type *yjbJ* sequence. These results prove that the YjbJ protein is nonessential under these environmental conditions and agree with the earlier results obtained by replacing the gene with the Km^r insertion allele.

Competition experiments to compare insertional and deletion phenotypes. To compare the phenotypes of the various mutant strains, isogenic strains with the insertional *yjbJ* and *hdeA* alleles were compared in a growth and survival competition with wild-type *E. coli*. In a second experiment, the *yjbJ* insertion and deletion strains were compared (see Materials and Methods). Figure 6A shows the survival of the *yjbJ* and the viable *hdeA* insertion mutants in competition with the wild-type strain. Under these conditions, the *hdeA* deletion causes a slight growth defect with respect to wild-type EMG2, while the *yjbJ* insertion strain outcompetes the wild-type strain. Figure 6B shows the competition results for the strain with the *yjbJ* insertional allele versus the strain with the deletion *yjbJ* allele. Surprisingly, two different phenotypes are observed for the different mutant alleles. In this assay, the *yjbJ* insertion strain outcompetes the *yjbJ* deletion strain.

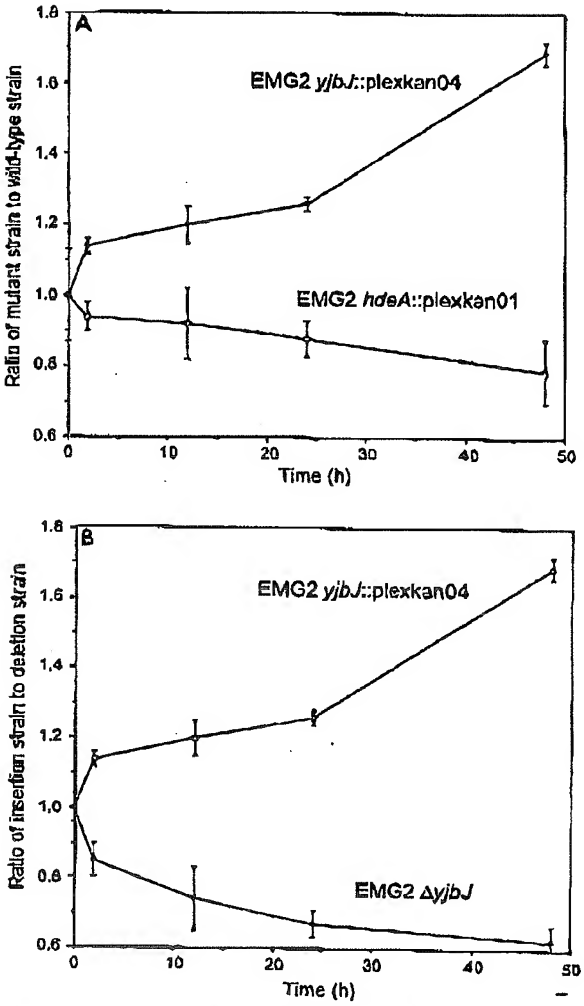

FIG. 6. Survival and growth competition between isogenic strains having either insertion or deletion alleles of *hdeA* and *yjbJ*. Equivalent numbers of cells from each strain were inoculated into rich medium and grown in competition under aerobic conditions at 37°C. At various time points, the cells were plated on rich media with and without kanamycin, and the viable cell density of each strain was assayed. (A) Relative survival of EMG2 *yjbJ*::plexkan04 and EMG2 *hdeA*::plexkan01 insertion strains when competed against wild-type EMG2. (B) Comparative survival of the EMG2 *yjbJ*::plexkan04 insertion strain and the EMG2 Δ*yjbJ* deletion strain when cultured together.

DISCUSSION

We have presented an improved method for performing gene replacements in *E. coli*. The method is similar to the pop-in/pop-out method used for *Saccharomyces cerevisiae* (6, 31, 33) and the hit-and-run procedure used for mouse embryonic stems cells (17). Unlike other methods used for gene replacements in *E. coli* that use ColE1 plasmids in a *polA1* background or transformation of linear DNA into *recBC*, *sbcB*, or *recD* strains, this protocol can be performed directly in wild-type strains (15, 21, 37, 44). Since the system is plasmid based, gene replacements are easily performed in any genetic background that is recombination proficient and supports the replication of pSC101 plasmids. Using this system, we have created another 44 *E. coli* strains with in-frame deletions of other ORFs (27a).

Although not attempted in our lab, the pKO3 gene replacement method can be used for constructing *E. coli* strains with multiple mutations without the need for multiple drug resistance markers or for replacing DNA sequences in the chromosome with precise point mutations. Finally, the method can be used for altering large exogenous fragments of DNA cloned into the single-copy P1 or BAC vectors which use *E. coli* as the host cell (38, 42).

In contrast to the deletion method, the insertion method creates mutations by inserting a Km^r gene (interposon) into cloned chromosomal DNA segments similar to a previous protocol (29). We designed this method for a gene that is predicted to be essential and uses selection instead of a screen to assess gene replacement. The Km^r gene was chosen as the marker since the gene has no homology to either the gene replacement vector pKO3 or the *E. coli* chromosome. We engineered the Km^r interposons with a different multiplex sequencing tag flanking each side of the interposon so that mutagenized clones could be sequenced by either cycle or multiplex sequencing (10, 27).

The two distinct phenotypes resulting from the insertional mutagenesis of *hdeA* highlight the unreliability of insertional mutagenesis. The comparison of the *yjbJ* insertion and deletion strains in the competition experiment also illustrates the phenotypic differences that can occur as a result of the particular type of mutation created. The *yjbJ* insertion strain appears to have an advantage over both the wild type and its respective deletion strain under the selection condition tested. Insertional mutagenesis has the potential for several undesired side effects, including polar termination-induced reduction of downstream operon expression (3), fusion products (2, 20), and misregulation of adjacent genes due to the insertion marker's promoter (11, 22, 43). Assigning a phenotype to a mutated gene may be problematic if the phenotype is actually a consequence of both the primary mutation and its effects on the surrounding genes.

This system, of replacing targeted ORFs with in-frame deletions was developed to reduce the inherent problems of insertional mutagenesis. Sensitive to the existence of transcrip-

tional and translational overlap in prokaryotic operons (13), our deletions were designed to retain translational coupling and to minimize the disruption of the regulation of neighboring genes in an operon (13, 24). The first six codons (18 bp) at the 5′ end of the gene were retained to maintain the gene's translation start signals. The last 12 codons (36 bp) at the 3′ end of the gene were retained based on the maximum overlap of coding regions observed in a sequence analysis of *E. coli* and *Salmonella typhimurium* operons (30). The largest observed overlap of a gene's 5′ coding region into a neighboring gene's 3′ coding region was 20 bp (*5′-cbiF-cbiG-3′* in the *cob* operon of *S. typhimurium*). The 36-bp overlap was chosen to maintain translational coupling in a gene cluster for operons with potentially even greater overlapping regions and for ambiguity in downstream translation start site assignment.

The expected frequency of colonies bearing the deletion allele after resolution of the plasmid integrates is 50%. As expected, the frequency of colonies bearing the Km^r insertion allele of *yjbJ* after plasmid resolution was approximately 50%. However, the observed resolution frequency of colonies with PCR-generated deletion alleles of both *yjbJ* and *hdeA* was only 3 to 7%. We speculate that this reduction in resolution frequency is caused by the reduced length of homologous sequences combined with possible PCR-generated DNA mismatches that flank one side of the duplications, causing the resolution of the integrate to be asymmetric (1, 34–36). In *E. coli*, a Chi site represented by the octanucleotide sequence 5′-GCTGGTGG-3′ stimulates recombination, depending on the length of the recombination interval and the location of the Chi site with respect to the interval (40, 41). We searched the genomic regions flanking *yjbJ* and *hdeA* and did not find the octanucleotide sequence in the vicinity of the two genes.

This research indicates that in the emerging post-genome sequencing era, when high-throughput evaluation of uncharacterized ORFs becomes a necessity, insertional mutagenesis by traditional methods will not be sufficiently reproducible to assign phenotypes based on subtle strain-by-strain variations. Because the engineering of in-frame deletions enables us to avoid many of the phenotypic artifacts mentioned earlier, we should be able to attach significance to a greater number of the phenotypes that we observe. This method will help investigators to systematically assign functions to the vast number of new ORFs revealed by current microbial sequencing projects.

ACKNOWLEDGMENTS

We thank Claire and Doug Berg for plasmid pBS-TS and the Kushner lab for plasmids pMAK700 and pMAK705. We thank Robert Kolter, Richard Baldarelli, Fritz Roth, and Pete Estep for helpful discussions. We are especially grateful to Elizabeth A. Malone and Martha Bulyk for critical reading of the manuscript.

This work was funded by DOE grant DE-FG02-87ER60565.

REFERENCES

1. Albertini, A. M., M. Hofer, M. P. Calos, and J. H. Miller. 1982. On the formation of spontaneous deletions: the importance of short sequence homologies in the generation of large deletions. Cell 29:319–328.
2. Bassford, P. J., T. J. Silhavy, and J. R. Beckwith. 1979. Use of gene fusion to study the secretion of maltose-binding protein into *Escherichia coli* periplasm. J. Bacteriol. 139:19–31.
2a. Berg, C. Personal communication.
3. Berg, C. M., and D. E. Berg. 1996. Transposable element tools for microbial genetics, p. 2588–2612. *In* F. C. Neidhardt, R. Curtiss III, C. Gross, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
4. Birnboim, H. C., and J. Doly. 1979. A rapid alkaline extraction procedure for screening recombinant plasmid DNA. Nucleic Acids Res. 7:1513.
5. Blomfield, I. C., V. Vaughn, R. F. Rest, and B. I. Eisenstein. 1991. Allelic exchange in *Escherichia coli* using the *Bacillus subtilis sacB* gene and a temperature-sensitive pSC101 replicon. Mol. Microbiol. 5:1447–1457.
6. Boeke, J. D., J. Trueheart, G. Natsoulis, and G. R. Fink. 1987. 5-Fluoroorotic acid as a selective agent in yeast molecular genetics. Methods Enzymol. 154:164–175.
7. Bollen, A., R. Lathe, A. Herzog, D. Denicourt, J. P. Lecocq, L. Desmurez, and R. Lavalle. 1979. A conditionally lethal mutation of *Escherichia coli* affecting the gene coding for ribosomal protein S2 (*rpsB*). J. Mol. Biol. 132:219–233.
8. Brown, E. D., E. I. Vivas, C. T. Walsh, and R. Kolter. 1995. MurA (MurZ), the enzyme that catalyzes the first committed step in peptidoglycan biosynthesis, is essential in *Escherichia coli*. J. Bacteriol. 177:4194–4197.
9. Chang, S.-Y. P., E. C. McGary, and S. Chang. 1989. Methionine aminopeptidase gene of *Escherichia coli* is essential for cell growth. J. Bacteriol. 171:4071–4072.
10. Church, G. M., and S. Kieffer-Higgins. 1988. Multiplex DNA sequencing. Science 240:185–188.
11. Ciampi, M. S., M. B. Schmid, and J. R. Roth. 1982. Transposon Tn*10* provides a promoter for transcription of adjacent sequences. Proc. Natl. Acad. Sci. USA 79:5016–5020.
12. Dedonder, R. 1966. Levansucrase from *Bacillus subtilis*. Methods Enzymol. 8:500–505.
13. Draper, D. E. 1996. Translational initiation, p. 902–908. *In* F. C. Neidhardt, R. Curtiss III, C. Gross, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. ASM Press, Washington, D.C.
14. Gay, P., D. Le Coq, M. Steinmetz, T. Berkelman, and C. I. Kado. 1985. Positive selection procedure for entrapment of insertion sequence elements in gram-negative bacteria. J. Bacteriol. 164:918–921.
15. Guttersen, N. I., and D. E. Koshland. 1983. Replacement and amplification of bacterial genes with sequences altered *in vitro*. Proc. Natl. Acad. Sci. USA 80:4894–4898.
16. Hamilton, C. M., M. Aldea, B. K. Washburn, P. Babitzke, and S. R. Kushner. 1989. New method for generating deletions and gene replacements in *Escherichia coli*. J. Bacteriol. 171:4617–4622.
17. Hasty, P., R. Ramirez-Solis, R. Krumlauf, and A. Bradley. 1991. Introduction of a subtle mutation into the *Hox-2.6* locus in embryonic stem cells. Nature 350:243–246.
18. Ho, S. N., H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. Gene 77:51–59.
19. Horton, R. M., H. D. Hunt, S. N. Ho, J. K. Pullen, and L. R. Pease. 1989. Engineering hybrid gene without the use of restriction enzymes: gene splicing by overlap extension. Gene 77:61–68.
20. Ito, K., P. J. Bassford, and J. R. Beckwith. 1981. Protein localization in *E. coli*: is there a common step in the secretion of periplasmic and outer-membrane proteins? Cell 24:707–717.
21. Jasin, M., and P. Schimmel. 1984. Deletion of an essential gene in *Escherichia coli* by site-specific recombination with linear DNA fragments. J. Bacteriol. 159:783–786.
22. Kendrick, K. E., and W. S. Reznikoff. 1988. Transposition of IS50L activates downstream genes. J. Bacteriol. 170:1965–1968.
23. Kleckner, N., J. Bender, and S. Gottesman. 1991. Uses of transposons with emphasis on Tn*10*. Methods Enzymol. 204:139–180.
24. Landick, R., C. L. Turnbough, and C. Yanofsky. 1996. Transcription attenuation, p. 1263–1286. *In* F. C. Neidhardt, R. Curtiss III, C. Gross, J. L. Ingraham, E. C. C. Lin, K. B. Low, B. Magasanik, M. Riley, M. Schaechter, and H. E. Umbarger (ed.), *Escherichia coli* and *Salmonella*: cellular and molecular biology, 2nd ed. Washington, D.C.
25. Link, A. J., K. Robison, and G. M. Church. 1997. Comparing the predicted and observed properties of proteins encoded in the genome of *Escherichia coli* K-12. Electrophoresis 18:1259–1313.
26. Maniatis, T., E. F. Fritsch, and J. Sambrook. 1989. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.
27. Murray, V. 1989. Improved double-stranded DNA sequencing using the linear polymerase chain reaction. Nucleic Acids Res. 17:8889.
27a. Phillips, D., and G. Church. Available at http://arep.med.harvard.edu/gmc/ecoko.html.
28. Ponce, M. R., and J. L. Micol. 1992. PCR amplification of long DNA fragments. Nucleic Acids Res. 20:623.
29. Prentki, P., and H. M. Krisch. 1984. *In vitro* insertional mutagenesis with a selectable marker. Gene 29:303–313.
30. Roth, J. R., J. G. Lawrence, M. Rubenfield, S. Kieffer-Higgins, and G. M. Church. 1993. Characterization of the cobalamin (vitamin B12) biosynthetic genes of *Salmonella typhimurium*. J. Bacteriol. 175:3303–3316.
31. Rothstein, R. 1991. Targeting, disruption, replacement, and allele rescue: integrative DNA transformation in yeast. Methods Enzymol. 194:281–301.
32. Rudd, K. E., W. Miller, C. Werner, J. Ostell, C. Tolstoshev, and S. G. Satterfield. 1991. Mapping sequenced *E. coli* genes by computer: software, strategies and examples. Nucleic Acids Res. 19:637–647.
33. Scherer, S., and R. W. Davis. 1979. Replacement of chromosome segments

with altered DNA sequences constructed *in vitro*. Proc. Natl. Acad. Sci. USA 76:4951–4955.
34. Senecoff, J. F., and M. M. Cox. 1986. Directionality in FLP protein-promoted site-specific recombination is mediated by DNA-DNA pairing. J. Biol. Chem. 261:7380–7386.
35. Shen, P., and H. V. Huang. 1986. Homologous recombination in *Escherichia coli*: dependence on substrate length and homology. Genetics 112:441–457.
36. Shen, P., and H. V. Huang. 1989. Effect of base pair mismatches on recombination via the RecBCD pathway. Mol. Gen. Genet. 218:358–360.
37. Shuvell, D. E., A. M. Abou-Zamzam, B. Demple, and G. C. Walker. 1988. Construction of an *Escherichia coli* K-12 *ada* deletion by gene replacement in a *recD* strain reveals a second methyltransferase that repairs alkylated DNA. J. Bacteriol. 170:3294–3296.
38. Shizuya, H., B. Birren, U.-J. Kim, V. Mancino, T. Slepak, Y. Tachiiri, and M. Simon. 1992. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc. Natl. Acad. Sci. USA 89:8794–8797.
39. Slater, S., and R. Maurer. 1993. Simple phagemid-based system for generating allele replacement in *Escherichia coli*. J. Bacteriol. 175:4260–4262.
40. Smith, G. R. 1987. Mechanism and control of homologous recombination in *Escherichia coli*. Annu. Rev. Genet. 21:179–201.
41. Smith, G. R. 1988. Homologous recombination in procaryotes. Microbiol. Rev. 52:1–28.
42. Sternberg, N. 1990. Bacteriophage P1 cloning system for the isolation, amplification, and recovery of DNA fragments as large as 100 kilobase pairs. Proc. Natl. Acad. Sci. USA 87:103–107.
43. Wang, A., and J. R. Roth. 1988. Activation of silent genes by transposons Tn*5* and Tn*10*. Genet. 120:875–885.
44. Winans, S. C., S. J. Elledge, J. H. Krueger, and G. C. Walker. 1985. Site-directed insertion and deletion mutagenesis with cloned fragments in *Escherichia coli*. J. Bacteriol. 161:1219–1221.
45. Yoshida, T., C. Ueguchi, and T. Mizuno. 1993. Physical map location of a set of *Escherichia coli* genes (*hde*) whose expression is affected by the nucleoid protein H-NS. J. Bacteriol. 175:7747–7748.
46. Yoshida, T., C. Ueguchi, H. Yamada, and T. Mizuno. 1993. Function of the *Escherichia coli* nucleoid protein, H-NS: molecular analysis of a subset of proteins whose expression is enhanced in a *hns* deletion mutant. Mol. Gen. Genet. 237:113–122.

Received Time May. 1. 4:38AM